

**Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft
Band 30**

**Evaluation von Subject Gateways des Internet
(EJECT)**

Andreas Kaizik
Winfried Gödert
Achim Oßwald

Juni 2001

Fachhochschule Köln
Fachbereich Bibliotheks- und Informationswesen

Kaizik, Andreas; Gödert, Winfried; Oßwald, Achim:

Evaluation von Subject Gateways des Internet
(EJECT) /

von Andreas Kaizik, Winfried Gödert, Achim Oßwald. -

Köln : Fachhochschule Köln, Fachbereich Bibliotheks- und
Informationswesen, 2001. -

(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 30)

ISSN (Print) 1434-1107

ISSN (elektronische Version) 1434-1115

*Die **Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft** berichten über aktuelle Forschungsergebnisse des Fachbereichs Bibliotheks- und Informationswesen der Fachhochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozent/inn/en, als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaftlichen Qualität der Veröffentlichungen liegt bei der Schriftleitung. Jeder Band erscheint parallel in Printversion und in elektronischer Version (über unsere Homepage: <http://www.fbi.fh-koeln.de/papers/index/titel/htm>).*

Fachhochschule Köln Fachbereich Bibliotheks- und Informationswesen Claudiusstr.1 D-50678 Köln
Tel.: 0221/8275-3376 Fax: 0221/3318583

Schriftleitung: Karin Finkbeiner, Prof. Dr. Wolfgang G. Stock

© by FH Köln 2001

Zusammenfassung

Der Umfang und die Heterogenität des Informationsangebotes erfordert immer differenzierte Methoden und Hilfsmittel für das gezielte und möglichst ballastfreie Auffinden von Informationsquellen im Kontext eines bestimmten Fachgebietes oder einer wissenschaftlichen Disziplin. Um dieses Ziel zu erreichen, wurden in der letzten Zeit eine Reihe sog. Subject Gateways entwickelt. Bislang liegen weder viele Untersuchungen zur Qualität derartiger Hilfsmittel vor noch ist eine differenzierte Methodik für solche Bewertungen entwickelt worden. Das Projekt Evaluation von Subject Gateways des Internet (EJECT) verfolgte daher die Ziele:

- durch Analyse bereits realisierter Subject Gateways die Verwendungsvielfalt des Begriffes aufzuzeigen und zu einer Präzisierung der Begriffsbildung beizutragen;
- einen methodischen Weg zur qualitativen Bewertung von Subject Gateways aufzuzeigen;
- diesen Weg anhand einer Evaluation des Subject Gateways EULER zu testen, das im Rahmen eines EU-Projektes für das Fachgebiet Mathematik entwickelt wurde.

Die Resultate der Evaluation werden in dieser Studie ausführlich vorgestellt und es wird aufgezeigt, inwieweit eine Übertragung auf die Bewertung anderer Gateways möglich ist.

Inhaltsverzeichnis:

<u>1. Fachliche Informationsrecherche im Internet</u>	5
<u>2. Das Kölner Projekt EJECT</u>	6
<u>3. Der Begriff Subject Gateway</u>	7
<u>4. Das EU-Projekt EULER</u>	10
<u>5. Die EULER-Evaluation</u>	14
<u>5.1 Die Ergebnisanzeige</u>	16
<u>5.2 Ziel der Untersuchung</u>	17
<u>5.3 Vorgehensweise</u>	18
<u>5.4 Suchfragen für die EULER-Evaluation</u>	19
<u>5.5 Erste Evaluationsphase: EULER und die Quelldatenbanken</u>	22
<u>5.6 Zweite Evaluationsphase: Aktualität</u>	24
<u>5.7 Dritte Evaluationsphase: Vergleich System-immanenter Suchstrategien</u>	24
<u>6. Evaluationsergebnisse</u>	25
<u>6.1 Generelle Auffälligkeiten</u>	25
<u>6.2 Retrievalfunktionalität</u>	26
<u>6.3 Aktualität der Datensätze</u>	31
<u>6.4 Immanente Evaluation</u>	33
<u>7. Ergebnistabellen</u>	35
<u>7.1 Tabellen zur Bewertung der Retrievalfunktionalität und Aktualität des Systems</u>	35
<u>7.2 Tabelle zur der immanenten Untersuchung</u>	59
<u>8. Schlussfolgerung</u>	61
<u>8.1 Zusammenfassende Bewertung der EULER-Retrievalévaluation</u>	61
<u>8.2 Evaluierung von Subject Gateways</u>	62

1. Fachliche Informationsrecherche im Internet

Der noch immer stetige Anstieg der über das Internet zugänglichen Informationen geht mit einer immer intensiveren Nutzung der Ressourcen auch für wissenschaftliche Zwecke einher. Dabei ist der enorme Datenbestand des Internets weit davon entfernt, umfassend und systematisch, im Sinne einer gezielten Auffindung relevanter Dokumente, geordnet zu sein. Insbesondere der Gedanke einer Bereitstellung nach bibliothekarischen Standards entspricht zur Zeit kaum der Realität und wird womöglich nie umfassend erreicht werden.

Daher führt eine Suche nach relevanten Informationen häufig zu Schwierigkeiten: Die oft noch primäre Nutzung der bekannten Suchmaschinen liefert im Sinne einer gezielten Suche häufig unbefriedigende Ergebnisse, da die meist hohen Trefferzahlen neben relevanten Dokumenten in der Regel weitaus mehr Dokumente aufweisen, die das gewünschte Thema nur unzureichend abdecken. Das Gros der gefundenen Dokumente setzt sich nicht selten aus eher populären als wissenschaftlich verwendbaren oder aber auch rein kommerziell orientierten Webseiten zusammen

Somit muss der Informationssuchende zunächst viele unterschiedliche Quellen kennen: Um diese schließlich effektiv nutzen zu können, ist es oft unerlässlich, zwischen einer wachsenden Anzahl von Systemen mit verschiedenen Suchmasken zu wechseln. Eine Gewöhnung an unterschiedliche Benutzeroberflächen sowie das Erlernen von mehreren Retrievalsprachen ist daher eine zeitraubende und lästige Notwendigkeit. Da oftmals nur mangelnde Kenntnisse der vielen unterschiedlichen Retrievalsysteme bestehen, können zudem nicht immer alle angebotenen Möglichkeiten der Systeme ausgenutzt werden. Eine kaum vorhandene Integration der bestehenden Angebote untereinander verringert die Effektivität zusätzlich. Zudem ist eine Lokalisierung der gesuchten Information, eben aufgrund des reichhaltigen Angebots, oftmals schlecht möglich.

Um die Umgebung des Internets den Bedürfnissen des für wissenschaftliche Zwecke Recherchierenden anzupassen und um die geschilderten Schwierigkeiten wenigstens zu vermindern, kam es in den letzten Jahren verstärkt zu der Einrichtung von Systemen, deren Ziel es ist, die Angebote bezogen auf die verschiedenen Fachgebiete zu strukturieren und somit eine Verbesserung der Recherchemöglichkeiten im Internet zu erreichen. Diese Entwicklung und die Errichtung von Portalen und den oft umfangreicheren Subject Gateways wird weiterhin zunehmend, oft in kooperativen Formen und durch entsprechende Fördermaßnahmen, forciert.

2. Das Kölner Projekt EJECT

Da sich Subject Gateways bereits heute zunehmend etabliert haben und bei der Suche nach wissenschaftlich relevanten Dokumenten nicht unberücksichtigt bleiben können, stellt sich die Frage, wie eine objektive Bewertung der verschiedenen Angebote durchgeführt werden kann. Verfahren, die methodisch an herkömmlichen Retrievaltests von klar begrenzten Datenbanken anknüpfen, sind in der bisherigen Form hier nicht anwendbar.

Um einen entsprechenden Ansatz zu finden, fand das an der Fachhochschule Köln im Fachbereich Bibliotheks- und Informationswesen (von Juli 2000 bis März 2001) durchgeführte Projekt EJECT (Evaluation von Subject Gateways des Internet) statt. Als Einstieg wurde in diesem Projekt zunächst eine Evaluation des kooperativen EU-Projektes EULER durchgeführt. Der Untersuchungsschwerpunkt lag dabei auf der Retrievalfunktionalität der EULER-Suchmaschine. In dieser Evaluation konnte die speziell entwickelte Methodik erstmals angewendet und erprobt werden. Die so gewonnenen Erkenntnisse sollen hier in generalisierter Form vorgestellt werden, um anzudeuten, wie sie auf die Evaluation anderer Subject Gateways übertragen werden kann.

Zuvor ist es jedoch notwendig, den Begriff des Subject Gateways näher zu erläutern, bei der Verwendung dieses Begriffs eine gewisse Bedeutungsvielfalt zu beobachten ist.¹

¹ Weiterführende Literatur:

Bertelmann, R., Rusch-Feja, D.: Informationsretrieval im Internet. In: Internet : Grundlagen, Technik, Anwendungen. 1997. S. 72-76

Campell, D.: Australian subject gateways: political and strategic issues. In: Online information review. 24(2000) no.1, S. 73-77.

Dempsey, L.: The subject gateway: experiences and issues based on the emergence of the Resource Discovery Network. In: Online information review. 24(2000) no.1, S. 8-23.

Fischer, T., Neuroth, H.: SSG-FI – special subject gateways to high quality Internet resources for scientific users. In: Online information review. 24(2000) no.1, S. 69-72.

Heery, R.: Information gateways: collaboration and content. In: Online information review. 24(2000) no.1, S. 40-45.

Hickey, T.R.: CORC: a system for gateway creation. In: Online information review. 24(2000) no.1, S. 49-53.

Koch, T.: Quality-controlled subject gateways: definitions, typologies, empirical overview. In: Online information review. 24(2000) no.1, S. 24-34

Lim, E.: Subject Gateways in Südostasien: Anwendungen von Klassifikationen.

<http://www.ifla.org/IV/ifla65/papers/011-117g.htm>

MacLeod, R.: Promoting a subject gateway: a case study from EEVL. In: Online information review. 24(2000) no.1, S. 59-63.

Peereboom, M.: DutchESS: Dutch Electronic Subject Service – a Dutch national collaborative effort. In: Online information review. 24(2000) no.1, S. 46-48.

Place, E.: Internationale Zusammenarbeit bei Internet Subject Gateways.

<http://www.ifla.org/IV/ifla65/papers/009-143g.htm>

Neuroth, H., Lepschy, P.: Das EU-Projekt Penardus. In: Bibliotheksdienst. 35 (2001) H.1, S. 49-63.

Price, A.: NOVAGate: a Nordic gateway to electronic resources in the forestry, veterinary and agricultural sciences. In: Online information review. 24(2000) no.1, S. 69-72.

3. Der Begriff Subject Gateway

Die Entwicklung des World Wide Web seit 1991, eröffnete die Möglichkeit, die wissenschaftliche Kommunikation intensiver und direkter zu gestalten und brachte es mit sich, dass auch erste Suchhilfen entstanden. Diese bestanden zunächst aus Linksammlungen und werden auch heute noch als *subject trees* angeboten, um über einen fachlichen Zugang die verzeichneten Webseiten mittels Browsing einzusehen. Eine Weiterentwicklung dieses Konzeptes, vor dem Hintergrund der rasanten Zunahme der Anzahl von Webseiten, sind die Subject Gateways. Die Grundlage des Angebots bildet eine intellektuelle Auswahl von fachspezifischen Web-Ressourcen. Der Nutzer erhält zudem zusätzliche Informationen anhand einer Beschreibung der verzeichneten Webseiten. Zumeist wird auch eine Suchfunktion angeboten, die auf den Beschreibungen der Webseiten aufsetzt: Damit unterscheidet sich die Vorgehensweise von der Volltextindexierung bei robotergestützten Suchmaschinen. Die ersten europäischen Projekte dieser Art starteten bereits 1994, z.B. das britische Projekt SOSIG (Social Science Information Gateway).

Zur gleichen Zeit entstanden auch die ersten globalen, fachübergreifenden Suchmaschinen mit automatischer Indexierung (z.B. Alta Vista) sowie Webkataloge (z.B. Yahoo!) mit einem Suchzugriff, der über einen systematischen Pfad realisiert wird. Die Zahl der Suchdienste im Internet ist seitdem stetig gestiegen und hat nun eine gewisse Sättigung erfahren.

Die generelle Leistungsfähigkeit, Größe, Deckungsbereich und Anzahl der Funktionen der Suchmaschinen hat seit ihren Anfängen stark zugenommen.¹ Trotz der genannten Verbesserungen bestehen weiterhin generelle Probleme bei der robotergenerierten Indexierung von Hypertextdokumenten durch Suchmaschinen, wie etwa das Auflösen des Kontextes und der Struktur der Dokumente oder auch die mangelnde Qualitätsauswahl des Inputs. Dabei steigert die zunehmende Konkurrenz unter den Anbietern um Werbeeinnahmen noch die Bemühungen, einen höheren Recall (anstelle zur Verbesserung der Precision) zu erzielen. Da die Rankingmechanismen zumeist nicht in ihren Details offen liegen, werden Unsicherheiten der Nutzer in Bezug auf die Aussagekraft der Trefferzahlen noch forciert.

Neuere Entwicklungen, wie etwa Google oder AskJeeves, stellen interessante Ansätze dar und erbringen verbesserte Ergebnisse, an der prinzipiellen Begrenzung einer qualitativ ungefilterten Quellengrundlage können jedoch auch sie nichts ändern. Für Recher-

¹ Die meisten indizierten Seiten, mit ca. 600 Mio., hat z.Zt. Google, dicht gefolgt von FAST mit ca. 575 Mio.

chen nachqualitätskontrollierten Angeboten in einem definierten fachlichen Kontext sind auch diese Werkzeuge nur bedingt geeignet.

Hierdurch rückt die unzureichende Nachweissituation des Internets verstärkt in das Zentrum der Betrachtung. Für qualitätsgeprüfte Angebote in einzelnen Wissenschaftsdisziplinen besteht daher weiterhin ein Bedarf an Suchhilfen und Nachweisinstrumenten.

Es wird im zunehmenden Maße erkannt, dass Suchende großen Nutzen aus einem Service ziehen können, der potentiell relevante Webressourcen auswählt, evaluiert und beschreibt. Diese Entwicklung bietet aber nicht nur eine Vielzahl von Möglichkeiten, wie etwa für die bibliothekarische Interneterschließung, sondern auch das Risiko, dass kommerzielle Anbieter in diesem Bereich eine lukrative Nische finden werden, was wiederum zur Folge haben könnte, dass eine größere Aufmerksamkeit auf die Selektion der zu verwendenden Dienste gelegt werden muss.

Bei der Recherche nach wissenschaftlich relevanten Quellen bieten die globalen Dienste wie Suchmaschinen und Webkataloge keine befriedigende Suchfunktionen und – ergebnisse. Da sich immer mehr Bereiche der wissenschaftlichen Kommunikation des Internets bedienen, verstärkt sich dieses Problem noch durch die wachsende Bedeutung für den Forschungs- und Bildungssektor.

Es besteht also ein zunehmender Bedarf an Instrumenten und Angeboten, die eine fachspezifische Suche im Internet erleichtern und effizienter gestalten. Subject Gateways wollen hier einen der fortschrittlichsten Lösungswege anbieten.

Geprägt wurde der Terminus „Subject Gateway“ oder auch „Subject-based Information Gateways“ durch die ersten britischen Projekte im Rahmen des eLib-Programms (Mitte der 90er Jahre). Die Zielsetzung der britischen Bemühungen war es, eine organisierte Datenbank mit einer Beschreibung von Internetquellen zu erstellen, in der der Nutzer „browsen“ und suchen kann. Durch einen Prozess der Identifikation, Auswahl, Beschreibung, Klassifizierung und Indexierung von Webressourcen sollte eine verbesserte Zugänglichkeit der Internetdokumente erreicht werden.

Um den Begriff Subject Gateway genauer zu konkretisieren, werden - nachdem nun vielfältige Formen der Grundidee als Angebote im Netz bereitstehen - verschiedene Eingrenzungsversuche unternommen. Dabei wird von einer Anzahl ähnlicher Projekte ausgegangen und versucht, Gemeinsamkeiten und kennzeichnende Merkmale zu defi-

nieren. Eine Abgrenzung des Begriffs fällt dabei, je nach Auswahl der betrachteten Projekte, unterschiedlich eng oder weit aus.

Folgt man Traugott Koch, so sind Subject Gateways „*Internet-services which support systematic resource discovery. They provide links to resources (documents, objects, sites or services), predominantly accessible via the Internet. The service is based on resource description. Browsing access to the resources via a subject structure is an important feature.*¹

Bereits mit dieser Definition werden nicht-fachlich strukturierte Linklisten ausgeklammert. Koch nennt ca. 50 derzeit bestehende Dienste, die seiner Definition entsprechen.² Dabei umfasst Koch's Auflistung die Bandbreite von Linksammlungen mit einer minimalen Beschreibung bis zum fachlichen Service mit einem hohen Standard für Qualitätskontrolle und einer reichhaltigen bibliographischen Beschreibung.

Schließlich setzt Koch von dem *weiten* Terminus der Subject Gateways den *engen* Terminus der Quality-controlled Subject Gateways³ ab:

*Quality-controlled subject gateways 'are Internet-services which apply a rich set of quality measures to support systematic resource discovery. Considerable manual effort is used to secure a selection of resources which meet quality criteria and to display a rich description of these resources with standards-based metadata. Regular [record] checking and updating ensure good collection management. A main goal is to provide a high quality of subject access through indexing resources using controlled vocabularies and by offering a deep classification structure for advanced searching and browsing.'*⁴

Das Subject Gateway impliziert also gewisse Qualitätskriterien wie die intellektuelle Arbeit zur Auswahl und Beschreibung der Ressourcen. Dies soll durch qualifiziertes, fachlich ausgebildetes Personal, etwa Fachreferenten oder andere Wissenschaftler (im Gegensatz zu einer automatischen Indexierung) geschehen.

¹ Koch, T.: Quality-controlled subject gateways, 2000, S. 24

² z.B.: Dutch Electronic Subject Service (DutchESS), Social Science Information Gateway (SOSIG), Deutscher Bildungsserver, MathGuide, NOVAGate, Nordic Gateway to Information in Forestry, Veterinary and Agricultural Sciences

³ In der übrigen Literatur entspricht diese Beschreibung des quality-controlled Subject Gateways eher dem Begriff des Subject Gateways, Subject-Information based Gateway oder auch Information Gateway. Diese weitaus üblichere Bezeichnung Subject Gateways wird auch im weiteren Verlauf der Arbeit benutzt.

⁴ Koch, T.: Quality-controlled subject gateways, 2000, S. 24

Einen weiteren Maßstab stellen die Qualitäts- bzw. Selektionskriterien dar, die zur Auswahl der Dokumente herangezogen werden. Diese können sich sowohl auf inhaltliche als auch auf formale und technische Aspekte beziehen und sollten möglichst offen dargestellt werden.

Die Quellenbeschreibung mit Hilfe von Metadaten bildet einen weiteren Punkt, wobei hier zunächst die Anzahl der verwendeten Kategorien von Bedeutung sind. Die Verwendung allgemeiner Standards, z.B. Dublin Core, ist für mögliche Kooperationen unerlässlich, um Datenaustausch und Integrationsmöglichkeiten zu schaffen.

Ein wichtiger, aber gleichzeitig aufwendiger Aspekt ist ein gut organisiertes Sammlungsmanagement mit der damit verbundenen Datenpflege. In Anbetracht des dynamischen und sich ständig wandelnden Mediums Internet, ist die Datenpflege von besonderer Bedeutung, da ständig neue Quellen hinzukommen, sich verändern oder auch wegfallen.

Schließlich soll der inhaltliche bzw. fachliche Zugang durch ein kontrolliertes Vokabular und/oder eine Klassifikation gewährleistet sein, zumal gerade die klassifikatorische Erschließung in der Regel die Grundlage für die angebotene Browsing-Funktion bildet.

Subject Gateways sind Hilfsmittel für die fachliche Informationssuche im Internet, - wobei die Summe der genannten Qualitätskriterien sie von den herkömmlichen Instrumenten wie Suchmaschinen und Webkatalogen abgrenzen. Sie bilden einen innovativen Ansatz für eine konsequente Interneterschließung auf der Basis bibliothekarischer Standards.

4. Das EU-Projekt EULER

Legt man die der vorausgegangenen Definition von Subject Gateways zu Grunde, handelt es sich bei EULER um einen Sonderfall. Die genannten Qualitätskriterien werden von EULER größtenteils erfüllt.¹ Gleichzeitig sind jedoch weitere Eigenschaften vorhanden, so dass sich die Gesamtstruktur EULER's von „anderen“ Subject Gateways grundsätzlich unterscheidet.

Das im April 1998 gestartete EULER-Projekt (**E**uropean **L**ibraries and **E**lectronic **R**esources in Mathematical Sciences) wurde durch die Europäische Union im Rahmen des Telematics for Libraries Programms gefördert. Es wurde von der European Mathematical Society (EMS), der nahezu alle europäischen mathematischen Fach-

¹ Eine entsprechende Einordnung EULER's erfolgt in einem späteren Kapitel.

gesellschaften angehören, initiiert. Die Zielsetzung bestand darin, einen integrierten Suchdienst für mathematisch interessierte Nutzer (vom Studenten bis zum Hochschullehrer oder in der freien Wirtschaft tätigen Mathematiker) zu entwickeln, der über das Internet zugänglich ist und dabei einen Zugriff auf alle mathematisch relevanten Informationsquellen erlaubt.

Die dem EULER-Konsortium angehörigen Organisationen werden im folgenden genannt; diese stellen nicht nur die mathematisch relevanten Ressourcen ihrer jeweiligen Institute zur Verfügung, sondern waren daneben in der Planung und Durchführung des Projektes beteiligt:

- FIZ Karlsruhe (Dept. Math. & Comput. Sci., Berlin)
- The European Mathematical Society
- Technische Universität Berlin
- Cellule de Coordination Documentaire Nationale pour les Mathématiques
- NetLab, Lund University Library
- SUB Göttingen
- Centrum voor Wiskunde en Informatica, Amsterdam
- Università degli Studi di Firenze

Zusätzlich wurden die Preprints von MathNet integriert. Die Koordination des Projektes oblag dem FIZ Karlsruhe.¹

Aus der Summe der Projektpartner und den von ihnen bereitgestellten Daten ergibt sich, dass mit EULER auf die Bestände 13 verschiedener Datenbanken zugegriffen werden kann, die eine große Bandbreite von relevanten Datenangeboten abdecken:

- Qualitätskontrollierte bibliographische Datenbanken mit dem Nachweis von unterschiedlichen Dokumenttypen (Zeitschriftenaufsätze, Reports, Monographien, etc.)
- Online-Bibliothekskataloge (OPACs) und Dokumentenlieferdienste
- elektronische Zeitschriften (im Volltext)
- Archive von Vorabdrucken (Preprints) und „grauer“ Literatur
- qualitätskontrollierte Zugänge zu Fachinformationen im Internet
- Roboter-generierte Indizes anderer relevanter Internetquellen

¹ Weiterführende Informationen unter: <http://www.emis.de/projects/EULER/partners.html>

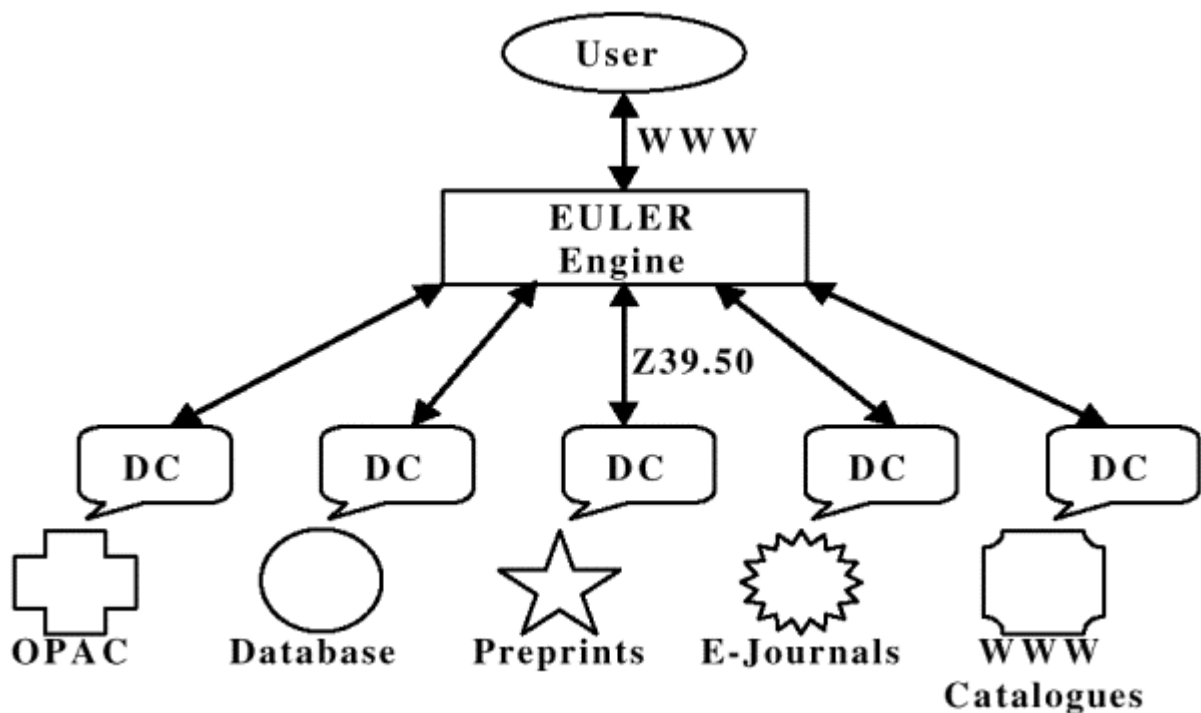
Sämtliche Partner bleiben während ihrer Teilnahme in Bezug auf ihre wissenschaftlichen und organisatorischen Entscheidungen autonom. Lediglich die Verwendung bestimmter Standards, Formate und Protokolle sind eine unabdingbare Voraussetzung für eine Teilnahme. Da EULER als offenes System konzipiert ist, wird erwartet, dass weitere Datenanbieter zu den bisherigen hinzukommen, um den Gesamtdatenbestand von derzeit ca. 900.000 Dokumenten noch zu vergrößern und den Abdeckungsbereich zu erweitern.

Das Hauptziel des Projektes ist es den anfangs geschilderten Schwierigkeiten bei der Informationssuche entgegenzuwirken und dabei einen integrierten Zugang auf heterogene Quellen verschiedener Art zu bieten. Somit ermöglicht es EULER, mit einer Suchmaske alle Quelldatenbanken anzusprechen und abzufragen. Die Einarbeitung in unterschiedliche Retrievalsprachen und –Systeme entfällt somit. Philosophie von EULER ist eine Dienstleistung im Sinne des One-Stop-Shops. Es soll möglich sein, von der Recherche und Lokalisierung relevanter Quellen bis hin zur Lieferung der gewünschten Dokumente (sowohl elektronisch als auch in Buch-/Zeitschriftenform über den eingegliederten Dokumentenlieferdienst) alle Vorgänge von einer Stelle aus zu initiieren.

Die Integration der Quelldatenbanken mit unterschiedlichen Strukturen in ein homogenes System wird durch eine Metadatenbeschreibung der jeweiligen Dokumente auf der Grundlage des Dublin Core Element Sets erreicht. Die Metadatenbeschreibungsätze sind, abhängig von der Art und Menge der Dokumente der Quelldatenanbieter, automatisch generiert, konvertiert oder mit Hilfe der Metadata Creator Software erstellt.

Zudem erlaubt die Verwendung des Z 39.50-Datenübertragungsprotokolls eine verteilte, parallele Abfrage auf den Datenbestand der teilnehmenden Anbieter.

Das Konzept wird nachstehend veranschaulicht:



Die Abbildung zeigt, dass der Nutzer mittels des HTTP Protokolls – genauer mit der Funktion GET - den Webserver anspricht, auf dem sich die EULER Engine befindet. Mittels eines CGI Programms wird dann die EULER-Engine angesprochen. Die Engine besteht aus mehreren Programmmodulen des Europagateways¹ und der speziellen Z 39.50-Variante Zebril². Mittels des YAZ Clients³ werden die verteilten Z 39.50 Server abgefragt.

Aufgrund dieser technischen Voraussetzungen ist es möglich, das EULER-System und dessen Kern, die EULER-Engine, über die herkömmlichen WWW-Browser zu nutzen. Für die Entwicklung der EULER-Suchmaschine wurde auf frei verfügbare Softwarekomponenten anderer EU-Projekte zurückgegriffen, um sie im Hinblick auf die speziellen Ziele EULER's zu modifizieren. Die so entwickelten Komponenten stehen für die Nachnutzung zukünftiger Projekte bereit.

¹ weitere Informationen unter: <http://europagate.dtv.DK/reports.htm>

² <http://hmm.dtv.DK/zebril/>.

³ <http://www.indexdata.dk/yaz/>

5. Die EULER-Evaluation

Im Rahmen des EULER-Arbeitsplanes wurden verschiedene Evaluationen durchgeführt, die zur Beurteilung des Systems dienen sollen. Neben der Untersuchung, die an der Fachhochschule Köln stattfand, die primär eine Überprüfung der Retrievalfunktionalität zum Ziel hatte, wurde am Fachbereich Bibliothek und Information der FH-Hamburg eine Usability-Studie durchgeführt und der von der SUB Göttingen eingegliederte Dokumentenlieferdienst des EULER-Systems bewertet.

Für die Durchführung der Evaluation war zunächst zu entscheiden, mit welcher Suchmaske gearbeitet werden sollte. Die EULER-Suchmaschine bietet zwei Suchmasken für die Recherche an: Die Simple Search beschränkt sich in ihren Möglichkeiten auf die am häufigsten benutzen Suchkriterien: „Alle Felder“, „Autor“, „Titel“ und die „MSC¹“. Eine Selektion auf einzelne Datenbanken des Systems ist hier nicht möglich. Es kann jeweils nur in einem der gewählten Felder recherchiert werden. Bei der Eingabe mehrerer Suchbegriffe werden diese als Phrase interpretiert, Boolesche Operatoren können nicht zum Einsatz kommen.

Mit der Full Search-Suchmaske werden die Recherchemöglichkeiten stark erweitert. In der Kölner Evaluation wurde ausschließlich mit dieser Suchmaske (der Beta-Version) gearbeitet. Diese Suchmaske ist in nachstehender Abbildung wieder gegeben und, und soll zunächst beschrieben werden:

¹ Mathematics Subject Classification

Abb. EULER Full Search Suchmaske

Die Full Search-Suchmaske erlaubt es, einzelne Quelldatenbanken separat auszuwählen oder auch bestimmte Arten von Datenbanken (z.B. Quality Controlled Resources, Bibliographic Services etc.) zu spezifizieren. Selbstverständlich kann hier aber auch auf eine Selektion verzichtet und so die gesamte Spannweite des Datenbestandes angesprochen werden.

Weitere Tools zur individuellen Suche bilden die folgenden Suchfelder, die nach Bedarf mit Booleschen Operatoren verknüpft werden können:

- All Fields
- Title
- Keywords
- LCSH
- MSC
- DDC

- CCS
- Date
- Description
- Event Date
- Event Location
- Event Name
- Format
- Physical Format
- Fulltext
- Item Identifier
- ISBN
- ISSN
- Language

Zudem ist eine Einschränkung nach Art der Dokumentenquellen möglich. Die unterschiedlichen „Resource types“ sind:

- Abstract
- Article
- Bibliography
- Executable
- Film
- Homepage
- Image
- Lecture Notes
- Monograph
- Num. Data
- Patent Specification
- Preprint
- Proceedings
- Reference
- Review
- Separatum
- Serial
- Software
- Source Code
- Tech. Report
- Text
- Thesis

Somit bietet das System mit den 19 unterschiedlichen Suchfeldern und mit der Unterscheidung von 22 Quellenarten ein außerordentlich umfangreiches Spektrum an Spezifizierungen für den Retrievalvorgang. Wieweit diese einsetzbar bzw. wie sinnvoll ihr Einsatz für konkrete Recherchen ist, war Gegenstand der Evaluation und wird später dargestellt.

5.1 Die Ergebnisanzeige

Die Anzeige der Treffer nach erfolgter Recherche erscheint in übersichtlicher, dublettenbereinigter Form. Im Kopf der Ergebnisanzeige wird neben der Gesamttrefferzahl zunächst die Summe der Treffer, gesplittet nach den angesprochenen Quelldatenbanken, angezeigt. Zudem liefert die Ergebnisseite Informationen über die Datensatznummer, den Autor, das Erscheinungsjahr, den Titel, Typ und über die Quelle des Dokuments. Alle Datensätze der Ergebnisseite werden nach Erscheinungsjahr in absteigender Reihenfolge aufgelistet. Außerdem besteht die Option,

alternative Darstellungsweisen nach Bedarf vorab einzustellen. Am Ende der Liste ist die vorhergegangene Eingabe in die Suchmaske nochmals mit einem zusätzlichen leeren Eingabefeld dargestellt, um eine weitere Verfeinerung der Suche zu ermöglichen.

5.2 Ziel der Untersuchung

Der Zweck aller Verfahren und Hilfsmittel zur Auffindung von Informationen gilt dem kontrollierten Umgang mit den jeweiligen Informationsressourcen. Einer der wichtigsten Aspekte ist das Auffinden der gesuchten relevanten Informationsquellen und der Zugriff darauf. Das Ziel der Untersuchung bestand darin, die Suchleistung des Retrievalsystems zu überprüfen und eine Bewertung im Hinblick auf etwaige Schwachstellen, vorzunehmen.

Die innovativen und sehr dynamischen Strukturen von „echten“ Subject Gateways - und hierfür ist das EULER-System ein gutes Beispiel - bringen es mit sich, dass Retrievaltests im herkömmlichen Sinne nicht angewendet werden können. Aus der Struktur des Systems ergibt sich, dass ständig neue Datensätze von den Quelldatenbanken zum bisherigen Bestand hinzukommen. Zugleich besteht die Möglichkeit (insbesondere im Hinblick auf die Webressourcen des Systems, dass andere wegfallen oder sich deren Inhalt ändert. Somit fehlen für herkömmliche Retrievaltests bestimmte Kenngrößen, die erst eine Bewertung ermöglichen. Es lässt sich z.B. nicht sagen, wie viele relevante Dokumente für eine bestimmte Abfrage im System enthalten sind. Eine Arbeit mit den Werten Recall und Precision, die wichtige Kriterien für die Bewertung der Retrievalfunktionalität bilden, ist für die Bestimmung des Recall unmöglich, für die Bestimmung der Precision zumindest stark eingeschränkt. Es lassen sich lediglich bedingte Aussagen über die Anzahl gefundener relevanter Datensätze erstellen, ebenso über die Anzahl der gefundenen nicht-relevanten Datensätze, die Vergleichsgröße der relevanten Datensätze, die nicht gefunden wurden, ist nicht zu ermitteln.

Wegen dieser Schwierigkeiten richten sich die angewendeten Testverfahren für derartige Systeme oft zunächst auf die Handhabbarkeit und Ergonomie. Es wird somit eine äußere Beurteilung der Suchmaske, der Summe und Art der Retrievalhilfsmittel, der Übersichtlichkeit, Navigation etc., oft mit Hilfe von unabhängigen Testpersonen, deren Recherchen dokumentiert werden, erzielt. Aussagen über den Inhalt sind nur insofern möglich, als dass der Gesamtdatenbestand im Verhältnis zum Bestand anderer Anbieter gesetzt wird. Weiterhin kann über die Qualität der Datensätze durch die Offenlegung des Auswahlverfahrens (hier dient insbesondere die Frage, ob die Datensätze eine Qualitätskontrolle durchlaufen, also intellektuell verifiziert werden) eine Bewertung gegeben werden, die häufig auch den Abdeckungsbereich mit einbezieht. Aber eine Be-

wertung der tatsächlichen Retrievalfunktionalität, im Hinblick auf die Auffindung der relevanten Dokumente, wird so nicht erreicht. Um dies zu ermöglichen wurde an der Fachhochschule Köln ein entsprechendes Konzept entwickelt und für die EULER-Evaluation angewendet.

5.3 Vorgehensweise

Die angewendete Methode wurde von der speziellen Struktur (hier insbesondere die Möglichkeit des gleichzeitigen Recherchierens der fachlich relevanten Datenbestände verschiedener Quelldatenbanken mittels **einer** Suchmaske) der Metasuchmaschine EULER vorgegeben.

Es wurde eine Liste von Suchanfragen für die Evaluation zusammen gestellt, die in ihrer Art möglichst den Suchfragen tatsächlicher Benutzer entsprechen. Daher musste zunächst ermittelt werden, an welche potentielle Nutzer sich EULER wendet. In den Planungspapieren steht hierzu die Aussage: Das Angebot richtet sich an alle Personen, bei denen ein Interesse an mathematischen Fachinformationen besteht und umfasst dabei das Spektrum vom Studenten bis zum Hochschullehrer oder dem in der freien Wirtschaft tätigen Mathematiker.

Auf dieser Grundlage wurden verschiedene Kategorien von Suchfragen gebildet, die eine möglichst breite Berücksichtigung aller Anfragearten erreichen sollte. Den unterschiedlichen Fragekategorien wurden schließlich, gestaffelt nach der voraussichtlichen Anfragehäufigkeit, eine entsprechende Menge von Einzelfragen zugeordnet. Dabei galt es zu berücksichtigen, dass in der Mathematik (wie in jedem anderen Fachgebiet auch) eine spezifische Semantik zum Tragen kommt, um unnötige Verzerrungen schon bei der Fragestellung zu vermeiden. Da die Datenbestände in EULER von mehreren europäischen Anbietern zur Verfügung gestellt werden, die mathematische Fachliteratur zum Großteil in Englisch verfasst wird und eine sprachliche Gleichförmigkeit erreicht werden sollte, wurden alle Eingaben in englischer Sprache formuliert. Insgesamt wurden 11 Kategorien mit insgesamt 83 Einzelfragen gebildet. Diese sind nachstehend wieder gegeben:

5.4 Suchfragen für die EULER-Evaluation

1. Life and works of mathematicians from earlier periods

Leopold Euler
Evariste Galois
Carl Friedrich Gauss
Emmy Noether
Gösta Mittag-Leffler

2. Materials on the work of contemporary mathematicians

Andrew Wiles
Gerd Faltings
Friedrich Hirzebruch
Helmut Neunzert

3. Nowadays problems in mathematics

1. Basic properties of double pullback graph transitions
2. Equivalence of Sobolev spaces
3. Asymptotic freeness almost everywhere for random matrices
4. Laser cooling and stochastics
5. Existence and uniqueness of weak solutions of an initial boundary value problem arising in laser dynamics
6. On the rareness of disturbed martingales in the class of all uniformly L^1 -bounded stochastic processes
7. Controlled algebraic K-theory
8. Two-primary algebraic K-theory of pointed spaces
9. Approximation by algebraic polynomials on rectangles
10. Schottky uniformizations of bordered compact Riemann surfaces
11. Anisotropic finite elements
12. The arithmetic and geometry of algebraic cycles
13. Wavelets in the geosciences
14. Calabi Yau manifolds, mirror symmetry
15. Knizhnik Zamolodchikov equation

4. Popular mathematical subjects

Fractals
Fermat theorem; Last Fermat theorem
Penrose tilings
Four color problem

Riemann Zetafunction

5. Teaching material to common university courses

Complex analysis
Linear Algebra
General Topology
Group theory
Statistics

6. Studies related informations/organisations

Masterstudiengang Mathematik, TU München
Postgraduate studies
Universität Kaiserslautern
EMS Summer School
CIMPA
CIME

7. Information on conferences

Internat. Congress of Mathematicians, Berlin: 18.-27. August 1998
1st international Congress of ISAAC (International Society for Analysis, Applications and Computing), University of Delaware, Newark
Mathematisches Forschungsinstitut Oberwolfach
17th International Symposium on Mathematical Programming, 7-11.08.2000
AMS Scand. 2000

8. References for books and journals, OPACs

Lugowski, Herbert/Weinert, Hanns J.: Grundzüge der Algebra
Pólya, György/ Szegő, Gábor: Aufgaben und Lehrsätze aus der Analysis
Riesz, Friedrich/Nagy, Bela: Vorlesungen über Funktionalanalyse
Demmel, James W.: Applied numerical linear algebra
Mathematica Scandinavica
Good, Phillip I.: Resampling Methods
Remmert, Reinhold: Funktionentheorie
Spiegel, Murray R.: Theory and problems of vector analysis and an introduction to tensor
Pentagon : a mathematics magazin for students
Studia mathematica

9. Mathematical Software

Cinderella
Mathematica
SPSS

10. Mathematical pictures, images etc.

Lorenz attractors
Kleinsche Flasche
Mandelbrot(menge)
Schneeflockenkurve

11. MSC-Notations and their verbalisations

1. Polyadic algebras	03 G 15
2. Umbral calculus	05 A 40
3. Weyl sums	11 L 15
4. Iwasawa theory	11 R 23
5. Cyclotomy	11 T 22
6. Excellent rings	13 F 40
7. Fano Varieties	14 J 45
8. Ulm sequences	20 K 12, 20 K 10
9. Hilbert algebras	46 K 15
10. Dual operator algebras	47 D 27, 47 L 45
11. Laguerre geometries	51 B 15
12. Steiner systems	51 E 10
13. Absolute planes	51 F 05
14. Kinematic spaces	51 J 15
15. Nearness spaces	54 E 17
16. Semimetric spaces	54 E 25
17. Moore spaces	54 E 30
18. Eckmann-Hilton duality	55 P 30
19. Homotopy functors	55 P 65
20. Pfaffian systems	58 A 17

Diese letzte Fragekategorie sollte Aufschluss darüber geben, ob die Verwendung von MSC-Notationen zu gleichen Ergebnissen führt, wie die verbale Suche. Als Beispiele wurden Sachverhalte gewählt, bei denen die Notation die gleiche Bedeutung und den gleichen Begriffsumfang aufweist wie die verbale Bezeichnung.

Auf der Grundlage der verschiedenen Kategorien und den jeweiligen Suchfragen wurde die Evaluation iterativ in drei Phasen durchgeführt, um Aussagen über die Retrievalfunktionalität, Aktualität und Zuverlässigkeit erstellen zu können.

5.5 Erste Evaluationsphase: EULER und die Quelldatenbanken

In der ersten Phase der Evaluation galt es, die tatsächliche Retrievalfunktionalität des EULER-Systems zu ermitteln. Eine Vergleichsgröße konnte erzielt werden, indem die über die EULER-Suchmaske eingegebenen Suchfragen zusammen mit den resultierenden Trefferanzeigen dokumentiert wurden, um sie anschließend den Ergebnissen einer separaten, direkten Recherche in den Quelldatenbanken gegenüberzustellen. In der Konzeption des Systems ist die Summe aller Datensätze der Quelldatenbanken identisch mit der Anzahl der Datensätze, die über EULER angesprochen werden können. Somit kann davon ausgegangen werden, dass die Ergebnisse einer Recherche über EULER (vorausgesetzt, die Suche wird auf alle teilnehmenden Datenbanken ausgedehnt) insgesamt der Summe der Ergebnissen einer separaten Abfrage der jeweiligen Quelldaten entspricht. Diesem Ansatz stehen jedoch einige Hindernisse entgegen:

Die von den Quelldatenbanken angebotenen Möglichkeiten zur Spezifizierung der Recherchen unterscheiden sich untereinander und von denen des EULER-Systems (hierdurch kann allerdings auch erkannt werden, ob die von EULER angebotenen Tools für mathematische Fachrecherchen zweckmäßig sind). Allerdings liegt in den voneinander abweichenden Spezifizierungsmöglichkeiten auch eines der schwerwiegendsten Probleme dieser Evaluation begründet: Wenn für dieselbe Anfrage in verschiedenen Datenbanken eine abweichende Suchsyntax verwendet wird, verzerrten sich die Ergebnisse bezüglich ihrer Vergleichbarkeit. Um diesem Problem entgegenzuwirken, wurde den Eingaben der Suchfragen das Axiom vorangestellt, dass zunächst die - in Anbetracht der zur Verfügung stehenden Tools - für die spezifische Anfrage sinnvollsten Hilfsmittel und Syntax zu verwenden seien. Kommen hierdurch unterschiedliche Anfrageformen zustande, ist die Recherche durch mehrfache Eingaben der Anfrage unter Verwendung anderer Tools zu ergänzen. Dies hat jedoch zur Folge, dass die Anzahl aller Eingaben nicht mehr nur der Menge aller (83) Einzelfragen - multipliziert mit den abzuarbeitenden Datenbanken - entspricht, sondern sich durch die unterschiedlichen Mehrfacheingaben erhöht.

Weiterhin galt es bei der Verwendung der von EULER angebotenen Retrievalhilfsmittel zu beachten, dass nicht für alle Hilfsmittel die entsprechenden Metadaten - mit einer hundertprozentigen Abdeckung - zur Verfügung stehen. Hier waren erhebliche Unterschiede vorzufinden: Beispielsweise sind bei der Benutzung der LCSH lediglich 1% der

Datensätze mit den entsprechenden Metadaten versehen, für die Felder Event Date, Event Location und Event Name sind es 3% usw.. Somit konnten Suchfelder, die aufgrund mangelnder Metadaten eine nur partielle Einbeziehung der vorhandenen Datensätze erlauben, niemals ausschließlich für eine Recherche herangezogen werden. Eine Verwendung dieser Suchfelder durfte daher lediglich als Ergänzung für eine vorausgegangene Recherche stattfinden.

Die Wahl der jeweils sinnvollsten Art einer Anfrageformulierung, unterliegt prinzipiell einer subjektiven Betrachtungsweise. Um dies möglichst auszuschließen, wurden alle entsprechenden Entscheidungen und Formulierungen nur von einer Person vorgenommen. Hierdurch wurden stetsso stets zumindest ähnliche, wenn auch subjektive, Maßstäbe angelegt.

Erst während der praktischen Arbeit stellte sich heraus, dass weitere Aspekte zu berücksichtigen sind, um homogene Anfragen zu erzielen. Eine möglichst große Kenntnis der verschiedenen Retrievalsysteme ist unerlässlich, damit die Ergebnisse der Recherchen in die Bewertung einfließen können. Somit ist vor Beginn jeder Recherche in den jeweiligen Quelldatenbanken eine gründliche Einarbeitung erforderlich. Diese ist nicht nur zeitraubend, sondern bleibt oft auch fragmentarisch, da die zur Verfügung stehenden Beschreibungen und Hilfetexte der Datenbanken häufig unzureichend erscheinen. Für die konkrete Untersuchung kam erschwerend hinzu, dass einige Anbieter sowohl ihre Beschreibungstexte, wie auch die Suchmaske selber, lediglich in der jeweiligen Landessprache zur Verfügung stellen. Daher ist es notwendig, dass vor der eigentlichen Eingabe der Anfragen zunächst Testrecherchen durchgeführt werden, um die folgenden Evaluationsrecherchen fehlerfrei durchführen zu können.

Um die Vergleichsgrößen möglichst umfassend zu ermitteln, wäre eine Einbeziehung tatsächlich aller Quelldatenbanken EULER's in die direkte, separate Abfrage wünschenswert gewesen. Einige der Quelldatenbanken konnten jedoch nicht einbezogen werden, da eine eigene Suchmaske fehlt, sie nur sehr beschränkte Tools anbieten oder nur ein alphabetischer Index vorhanden ist. Auch musste geprüft werden, wann eine Einbeziehung sinnvoll erscheint, da der Zeitraum der Evaluation drei Monate nicht übersteigen durfte. Daher konnten Datenbanken mit nur einem sehr geringen Umfang nicht berücksichtigt werden, da die Anzahl der Datensätze, etwa der ca. 1700 Dokumente der CWI-Reports im Verhältnis zu den ca. 900.000 Datensätzen im Gesamtsystem EULER als zu marginal zu werten war. Der zeitliche Aufwand hätte somit kaum im Verhältnis zu den Resultaten gestanden.

Diese Vorgehensweise war möglich, da die Ergebnisanzeige EULER's eine Auflistung der Trefferanzahlen separiert nach den jeweiligen Quellen dokumentiert. Daher konnten Treffer aus Quellen, die nicht separat abgefragt wurden, schlichtweg ignoriert werden und führten somit zu keiner Verfälschung der Untersuchung.

Für den Vergleich der Recherchen in EULER und den Quelldatenbanken wurde zunächst die reine Anzahl, der Recall, gefundener Treffer gewertet. Aus den aufgeführten Gründen konnten Maßstäbe wie Relevance und Precision nur sehr partiell in die Evaluation einfließen. Um aber neben der Trefferzahl weitere Bewertungsmerkmale zu finden, wurde in der Evaluation neben der Gesamttrefferzahl (wenn möglich) eine weitere Größe eingeführt, nämlich die Gesamttreffer abzüglich der eindeutig nichtrelevanten Dokumente.

Zudem sollten in der ersten Phase der Evaluation, neben den Retrievalergebnissen, weitere Auffälligkeiten des Systems dokumentiert und hinterfragt werden. Führten diese zu Fragen der Usability, so wurden sie nur dann Gegenstand der Evaluation, wenn sie im direkten Zusammenhang mit Rechercheergebnissen standen.

5.6 Zweite Evaluationsphase: Aktualität

Ein wichtiger Bestandteil des EULER-Systems sind die recherchierbaren Internet-Ressourcen (zum Teil qualitätskontrolliert). Diese erhöhen nicht nur den Abdeckungsbereich, sondern erheben daneben den Anspruch, über aktuelle Entwicklungen Informationen zu liefern. Daher sollte in der zweiten Phase die Aktualität EULER's überprüft werden. Einige der entwickelten Fragekategorien sind speziell auf eine aktuelle Thematik ausgerichtet (z.B. Nowadays problems in mathematics, Information on conferences).

Diese wurden ebenfalls mit geeigneter Syntax in die EULER-Suchmaske eingegeben sowie mittels der in Frage kommenden Quelldatenbanken (soweit recherchierbar) verglichen. Ein zusätzlicher Vergleich wurde durch eine zusätzliche Recherche mit der Meta-suchmaschine Metacrawler durchgeführt.

5.7 Dritte Evaluationsphase: Vergleich System-immanenter Suchstrategien

Die letzte Phase der Evaluation bildet einen Sonderfall: In der Mathematik hat die systematische Recherche mittels Klassifikation eine besondere Bedeutung. Die fachspezifische Klassifikation MSC wird häufig für die Informationssuche herangezogen und somit auch als Hilfsmittel von EULER angeboten. Ein Vergleich zweier Suchstrategien innerhalb desselben Systems ist demnach nahe liegend. Dafür war es zunächst erforderlich, MSC-Notationen zu ermitteln, die sich in einer klaren, eindeutigen und mög-

lichst kurzen Verbalisierung widerspiegeln. Auf dieser Basis wurden 20 Notationen mit den entsprechenden Verbalisierungen zusammengestellt (Fragekategorie 11). Die Notationen wurden in die Suchmaske eingegeben, die Ergebnisse dokumentiert. Für die Eingabe der Verbalisierungen konnte nur das Suchfeld „all fields“¹ in Frage kommen, um so einen Zugriff auf sämtliche Daten zu erreichen.

Da nicht alle Dokumente EULER's nach der MSC klassifiziert sind, bestand die Erwartung, dass eine Recherche mit den Verbalisierungen zu höheren Trefferzahlen führen sollte. Einen weiteren Vergleich ermöglichte auch hier eine verbale Recherche im Internet, mittels Metacrawler.

6. Evaluationsergebnisse

6.1 Generelle Auffälligkeiten

Mittels der EULER-Engine kann auf die Datensätze 13 verschiedener Datenbanken zugegriffen werden. Die über das Z 39.50-Protokoll verteilte Suche – ohne einen gemeinsamen Index - spricht jede der Datenbanken separat an. Aus unterschiedlichen technischen Gründen ist es möglich, dass eine oder auch mehrere der Quelldatenbanken nicht immer für die Recherche zur Verfügung stehen. Somit müssen Recherchen - insbesondere dann, wenn es erwünscht ist, dass eine bestimmte Datenbank in die Suche einbezogen wird - unter Umständen wiederholt werden. Der Nutzer erhält jedoch keine Angaben darüber, wann eine spezielle Datenbank wieder zur Verfügung steht.

Die Benutzeroberfläche (des „full search Modus“) erscheint teilweise wenig übersichtlich. Ungeübte und neue Anwender können aufgrund einiger Uneindeutigkeiten leicht den Überblick bezüglich der Auswahl der Datenbanken verlieren, sofern nicht sämtliche Quellen abgesucht werden sollen und besonders bei einer Selektion nach den „Resource Types“. In beiden Fällen ist die getroffene Auswahl schwer ersichtlich, da die Mehrzahl der Selektionsmöglichkeiten lediglich durch das Abscrollen in den jeweiligen Feldern erkennbar wird.

Viele der angebotenen Retrieval-Tools erscheinen sehr benutzerfreundlich und für eine effektive Recherche hilfreich. Leider sind manche davon - wie bereits angesprochen - nur bedingt einsetzbar. Das Fehlen entsprechender Metadaten führt zu einer oft nur fragmentarischen Recherche, die, abhängig vom verwendeten Tool, einen unterschiedlich großen Datenbestand unberücksichtigt lässt. Für den Benutzer besteht zwar die Möglichkeit, sich über die hinterlegten Metadaten der jeweiligen Retrievaltools zu infor-

¹ Zusätzlich wurden die Verbalisierungen dennoch auch als „Keywords“ eingegeben.

mieren, dies ist jedoch mit weiteren Arbeitsschritten verbunden und demnach nicht auf den ersten Blick ersichtlich. Somit besteht die Gefahr, dass ein Nutzer, der ein Tool gewählt hat, mit dem lediglich auf einen Teildatenbestand zugegriffen werden kann, nach erfolgloser oder unbefriedigender Suche glaubt, die entsprechenden Dokumente seien im System nicht vorhanden. Dies wird zudem dadurch verstärkt, dass einige der fraglichen Tools eine hohe Präzision suggerieren. Wird etwa nach einem Konferenzband zu einer bestimmten Tagung gesucht, bietet sich eine Verbindung von „Event Name“ und „Proceeding“ geradezu an, da so jeglicher Ballast ausgeschlossen sein sollte und das gewünschte Dokument unmittelbar aufgefunden werden müsste. Jedoch kann mit dieser Verknüpfung lediglich auf ein Prozent des Datenbestandes zugegriffen werden, was dazu führt, dass eine solche Recherche zumeist erfolglos bleibt.

Weiterhin musste festgestellt werden, dass es den Beschreibungstexten zur Erläuterung der Retrievalhilfsmittel an Eindeutigkeit mangelt. Aus diesen wird z.B. nicht ersichtlich, dass das Feld „all fields“¹ eben nicht, wie es der Name impliziert, alle Felder berücksichtigt. Für die professionelle Recherche ist es unabdingbar, dass die tatsächliche Bedeutung der Spezifizierungsmöglichkeiten offenkundig ist, da es ansonsten zu Verzerrungen der Rechercheergebnisse kommen kann.

Ein weiteres Problem zeigte sich bei der Verwendung von Trunkierungen. Hier kam es während der Evaluation wiederholt zu der paradoxen Situation (wie das unten aufgeführte Beispiel zeigt), dass eine Vergrößerung des Trunkierungsbereichs (bei dem selben Suchbegriff) zu einer Verringerung (sic) der Trefferzahlen führte.²

all fields=finite element*	4258 hits
all fields=finite elem*	4258 hits
all fields=finite ele*	3929 hits
all fields=finite el*	3915 hits
all fields=finite e*	3062 hits

6.2 Retrievalfunktionalität

Eine Abfrage mittels der EULER-Engine umfasst (sofern der „All Databases“-Modus gewählt wird) den selben Datenbestand wie die Summe der separaten, direkten Recherchen in den Quelldatenbanken.³ Somit muss im Optimalfall eine Übereinstimmung der

¹ Das Feld wurde mittlerweile in „basic index“ umbenannt.

² Das dargestellte Problem konnte mittlerweile behoben werden.

³ Dies gilt dann, wenn neu hinzugekommene Datensätze der Quelldatenbanken unmittelbar in das EULER-System überspielt werden. Daher war es notwendig, über alle Änderungen des Datenbestandes (d.h. über noch nicht stattgefundenen bzw. bevorstehende Datenupdates) seitens der EULER-Projektpartner informiert zu werden.

Rechercheergebnisse EULER's mit den separaten Abfragen in den Quelldatenbanken gegeben sein. Dieser Umstand bildete die Grundlage für die Überprüfung der Retrievalfunktionalität des EULER-Systems. Alle Abweichungen, insbesondere größerer Art, implizieren eine Ametrie des Retrievalverhaltens der Systeme.

Dieses Missverhältnis fiel bei der Untersuchung größer aus, als erwartet. Wie aus den in Kapitel 7 aufgeführten Tabellen, die eine Manifestation der unterschiedlichen Rechercheergebnisse vorweisen, ersichtlich wird, führte eine direkte Abfrage der Quelldatenbanken in vielen Fällen zu quantitativ höheren und auch zu qualitativ besseren Ergebnissen. Dies wurde bereits bei der ersten Fragekategorie (Life and works of mathematicians from earlier periods) erkennbar:

Bei der Recherche nach Dokumenten über Leben und Werk des Mathematikers Carl Friedrich Gauss wurde u.a. die Anfrageform 'MSC=01 a 70¹ & Titel=Gauss' gewählt. Eine Eingabe in EULER erbrachte insgesamt 36, davon 15 relevante Treffer. Von diesen 36 Treffern kamen 24 aus dem Zentralblatt MATH, eine direkte Recherche im Zentralblatt (mit der gleichen Anfrageform) ergab jedoch 69 Treffer, wovon 50 als relevant eingestuft wurden. Das heißt, tatsächlich im Zentralblatt vorhandene Datensätze konnten von der EULER-Engine nicht ermittelt werden. Zudem ist das Verhältnis von Recall und Precision bei der EULER-Recherche weitaus ungünstiger als bei der direkten Abfrage. Werden die Gesamttreffer der EULER-Recherche zugrunde gelegt, so ergibt sich eine Ballastrate von nahezu 60 %. Die separate Abfrage im Zentralblatt dagegen erbringt präzisere Ergebnisse, der Ballastanteil beträgt hier lediglich 28 %.

Dabei sind diese Resultate kein Ausnahmefall; in der gleichen Fragekategorie wurde nach Dokumenten zum Leben und Werk Leopold Euler's gesucht. Bei der Verwendung der MSC (auch hier 01 A 70) in Verbindung mit Titel = Euler findet die EULER-Suchmaschine 14 Treffer, davon 10 relevante. Von den 14 Treffern wurden 9 im Zentralblatt geortet, mit der direkten Abfrage konnten jedoch 41 Treffer (davon 26 relevante) nachgewiesen werden. Hier fällt zwar die Ballastrate weniger ungünstig aus, dennoch wurden 19 Dokumente aus dem Bestand des Zentralblattes von EULER nicht aufgefunden.

In der zweiten Fragekategorie (Materials on the work of contemporary mathematicians) war die Divergenz des Retrievalverhaltens weniger eindeutig. Oftmals entsprachen hier die über EULER ermittelten Treffer denen der direkten Abfragen. Aber auch hier konnten Unstimmigkeiten festgestellt werden:

¹ MSC: 01 A 70 = Biographies, obituaries, personalia, bibliographies

Im Rahmen der Aufgabenstellung wurde nach Gerd Faltings recherchiert. Für die EULER-Suchmaske wurde eine Oder-Verknüpfung von Keyword und Titel, jeweils mit dem Namen des Gesuchten, gewählt. Die Anfrage führte zu 26 insgesamt aber nicht-relevanten, Treffern. Von den 26 gefundenen Dokumenten kamen 4 von NetLab¹ (d.h. aus dem Internet) und 17 vom Zentralblatt. Weitere 5 kamen von Datenbanken, die hier nicht in die direkte Abfrage einbezogen wurden. Eine Recherche mittels Metacrawler, als Vergleichsgröße für die über NetLab erzeugten Treffer erbrachte 35 Dokumente. Die Recherche im Zentralblatt führte zu 33 mit 2 relevanten Treffern; hier wurden also beinahe doppelt so viele Dokumente aufgefunden wie mit EULER. Mit der EULER-Recherche wurde kein Treffer im Datenbestand der SUB Göttingen nachgewiesen, die direkte Abfrage führte jedoch zu 10 (wenn auch komplett nicht-relevanten) Treffern. Dabei gilt es zudem zu berücksichtigen, dass im Gegensatz zu der für die SUB und dem Zentralblatt angewendeten Suchmethode, beide Male lediglich Titel = Faltings, in EULER eine erweiternde Form, nämlich Keyword oder Titel = Faltings, gewählt wurde. Eine direkte Abfrage der Datenbestände von MathGuide, UNIFI, und den OPAC`s von Orsay und Strasbourg ergab, entsprechend den EULER-Ergebnissen, keine Treffer.

Unter den Fragen der vierten Kategorie (Popular mathematical subjects) finden sich bereits wieder größere Uneinheitlichkeiten:

EULER konnte bei der Suche nach dem Fermatschen Theorem 9 Dokumente auffinden (Eingabe: All fields = fermat theorem). Von diesen Treffern kamen 6 von NetLab, 2 vom Zentralblatt und einer vom CWI. Die unmittelbaren Recherchen in diesen drei Quelldatenbanken (anstelle von NetLab wurde auch hier Metacrawler herangezogen, da NetLab zum Zeitpunkt der Evaluation keine eigene Recherchemaske besaß) erbrachten mit 30, 31 und 21² weitaus größere Trefferquoten. Zudem konnten durch die direkten Recherchen weitere 29³ Dokumente aus anderen Datenbanken ermittelt werden, während EULER für diese keine Treffer anzeigte.

Außergewöhnlich starke Abweichungen zeigten sich bei der Suche zur Riemannschen Zetafunktion in der gleichen Kategorie:

In der EULER-Suchmaske wurde der Begriff als Phrase in trunkierter Form (riemann* zeta*) eingegeben. Es wurden 158 Treffer erzeugt. Obwohl die Resultate der Recher-

¹ Eine Einstufung nach Relevanz war bei diesen Treffern nicht möglich, ebenso wenig bei der Vergleichsrecherche mittels Metacrawler.

² Bei den Recherche im CWI-OPAC wurde dabei sogar eine einengendere Suchform, Titel = fermat theorem, benutzt.

³ MathGuide 2, SUB 15, UNIFI 1, Orsay/Strasbourg 11

chen über EULER und mittels der direkten Abfragen insgesamt noch recht gleichförmig ausfielen, zeigte sich in einem Fall eine völlig unverständliche Abweichung. Über EULER konnten 85 Dokumente aus dem Zentralblatt nachgewiesen werden. Die in der Syntax identische Recherche im Zentralblatt selber erbrachte jedoch 579 Treffer, also eine um beinahe 700 % (!) höhere Trefferquote!

Kurioserweise zeigte sich bei der Recherche zur Linearen Algebra (in der 5. Fragekategorie) ein fast gegensätzliches Phänomen: Mit EULER konnten insgesamt 126 Dokumente ermittelt werden, davon 125 aus dem Zentralblatt. Die direkte Abfrage führte aber lediglich zu 51 Treffern.¹ Allerdings zeigte sich bei der direkten Recherche in weiteren Quelldatenbanken, dass hier entsprechende Datensätze zu finden sind, die über EULER nicht aufgespürt wurden. Insgesamt setzt sich der allgemeine Trend in der 5. Fragekategorie fort: Die direkten Anfragen in den Quelldatenbanken führten zu quantitativ höheren Trefferzahlen. Auch wenn hier das Ziel der homonymen Frageformulierung seltener erreicht wurde, läßt die sehr große Divergenz bezüglich der Trefferzahlen darauf schließen, daß auch hier in den Quelldatenbanken vorhandene Dokumente von EULER nicht lokalisiert werden konnten. Dies wird auch dadurch deutlich, daß hier mehrere der EULER-Recherchen zu Null-Treffermengen führten, während die direkte Abfrage oftmals sehr hohe Treffermengen erbrachten.

Bei der Dokumentation der Abfragen zur sechsten Kategorie (Studies related informations/organisations) zeigte sich wiederum eine Divergenz der Retrievalergebnisse. Hier insbesondere im Hinblick auf die über EULER zugänglichen Internetressourcen. Eine Recherche zur Cimpa Summer School erbrachte über EULER lediglich drei Treffer von NetLab, die entsprechende Metacrawler-Recherche erbrachte statt dessen 13 Treffer, während ein gegensätzliches Verhältnis bei der Suche nach Informationen zu „Postgraduate Studies“ der Fall war. Hier konnten von EULER 114 Treffer von NetLab angezeigt werden, die Recherche im Internet selber erbrachte aber mit 25 nur ein weit-aus geringeres Ergebnis.

Insgesamt führten die Resultate, bezüglich der Internetressourcen zu sehr wechselhaften Rechercheerfolgen, wobei aber in den häufigeren Fällen die Recherche mittels Metacrawler der webrelevanten Suche mit EULER überlegen war.

¹ Allerdings wurde für die Recherchen eine leicht abweichende Retrievalsyntax verwendet: Die Euler-Eingabe war `Keyword=Course|Textbook&Keyword=Linear Algebra`; während im Zentralblatt die Eingabe nicht auf die Keywords beruhte, sondern anstelle dessen auf den Titel. Es darf aber davon ausgegangen werden, dass jedes Dokument, das im Titel Linear Algebra enthält auch unter diesem Keyword zu ermitteln ist. Umgekehrt ist es jedoch möglich, dass für ein Dokument zwar das Keyword vergeben wurde, ohne dass es im Titel enthalten ist. D.h., die Suchform nach Titel ist eine engere als nach Keyword.

Auch die 7. Kategorie (Information on conferences) bildete keine Ausnahme. Mit keiner der EULER-Recherchen konnten eine mit den direkten Abfragen übereinstimmende Ergebnisse erzielt werden. Hier waren die separaten Abfragen quantitativ wie qualitativ den EULER-Recherchen überlegen.

Die achte Fragekategorie (References for books and journals, OPACs) implizierte optimale Bedingungen für einen Vergleich der Retrievalfunktionalität der Systeme. Denn hier konnten sämtlich Eingaben in der EULER-Suchmaske in homogener Form auf die Suchmasken der Quelldatenbanken übertragen werden. Denn zur Ermittlung der entsprechenden Titel ist es stets ausreichend, lediglich mit den Suchfeldern Titel und/oder Autor zu arbeiten. Diese Felder werden sowohl von EULER als auch von allen in den Vergleich einbezogenen Datenbanken angeboten. Durch die Eindeutigkeit bei der Zuordnung der Suchergebnisse nach Relevanz kann es auch hierdurch kaum zu Verzerrungen kommen.¹ Daher ist die Aussagekraft der Ergebnisse der achten Fragekategorie besonders hoch zu werten.²

Die Resultate der Abfragen zur achten Kategorie zeigten insgesamt nur kleinere Abweichungen, aber auch hier fielen die Differenzen der Trefferzahlen (wenn auch geringfügiger) zu Gunsten der direkten Recherchen aus. In keinem der unter 8 eingegebenen Fälle war eine Suche über EULER erfolgreicher, als die separaten Abfragen. Umgekehrt konnten mit den direkten Recherchen weitere Bestandsnachweise ermittelt werden, die vom EULER-System unentdeckt blieben. Die höheren Trefferzahlen über die Quelldatenbanken sind einerseits zahlenmäßig gering, andererseits macht eine Erhöhung von einem auf zwei Nachweisen eine Differenz von 100 %. Somit darf auch hier gesagt werden, dass EULER hinter den Erwartungen zurück bleibt.

Die Recherchen zur 9. Fragekategorie (Mathematical Software) führten ebenfalls zu Divergenzen. In einem Fall wurde dies besonders deutlich: Gesucht wurde nach Dokumenten zur Software SPSS. Mit EULER (Frageformulierung: all fields=spss & keyword=software) wurden 6 Treffer – davon 5 relevante – ermittelt. Von diesen Treffern kam keiner von der SUB Göttingen, während eine Recherche im OPAC der SUB (Frageformulierung RSWK=SPSS) zu 45, durchweg relevanten, Treffern führte. Auch wenn hier eine abweichende Eingabeform vorliegt (eine Spezifizierung mittels RSWK wird von EULER nicht angeboten) verliert dieses Ergebnis kaum an Eindeutigkeit. Denn

¹ Jeder Nachweis wurde, nach Abzug eventueller Dubletten, als relevant gewertet.

² Hier muss jedoch darauf hingewiesen werden, dass die Treffer in den OPAC's von Orsay und Strasbourg von EULER zusammengefasst angezeigt werden. D.h., werden von EULER zwei Treffer angezeigt, bei der direkten Abfrage aber jeweils einer nachgewiesen, ist dennoch eine Übereinstimmung gegeben, sofern in den Beständen der OPAC's nicht die gleiche Ausgabe ermittelt wurde.

eine stichprobenartige Überprüfung der im OPAC der SUB georteten Dokumente hat ergeben, dass viele dieser Treffer bereits im Titel die Wörter SPSS und Software enthalten. Somit hätte EULER, bei der verwendeten Frageformulierung, zumindest diese auffinden müssen.

Stellt man den Gesamttrefferzahlen EULER's (in der 7. Kategorie) die Summe der Treffer der direkten Recherchen entgegen, führt dies zu ernüchternden Ergebnissen:

Mit EULER wurden bei der Recherche zur Software Mathematica insgesamt 23 Dokumente aufgefunden, die Summe der direkten Abfragen erbrachte 171 Treffer; also über 700% mehr.

Bei der Suche zu der SPSS fällt dies noch deutlicher aus: hier konnten anstelle von 6 Treffern 69 ermittelt werden.

Bei der Recherche nach der Software Cinderella ist das Mißverhältnis etwas abgeschwächt: Hier können den 28 EULER-Treffern 47 der direkten Abfragen entgegengestellt werden, was noch immer ein um 60% höheres Ergebnis ausmacht.

Die Recherchen zu der 10. Kategorie zeigten unterschiedliche Resultate. Von nahezu übereinstimmenden Trefferzahlen, eine Überlegenheit der direkten Abfragen bis zu vorteilhaften Ergebnissen der EULER-Recherchen war hier jede Variante möglich. Somit kann hier keine eindeutige Aussage zum Retrievalverhalten der Systeme erbracht werden. Um aussagekräftigere Resultate zu erzielen, müßten hier weitere Recherchen mit zusätzlichen Fragen abgearbeitet werden. Da aber Anfragen dieser Art (Bilder, Visualisierungen von mathematischen Problemen) von einem potentiellen Nutzer des EULER-Systems in der Realität wohl eher selten zu erwarten sind, wurde auf eine Ausweitung dieser Fragekategorie verzichtet; zumal die allgemeine Tendenz des Retrievalverhaltens anhand der anderen Kategorien ausreichend belegt erscheint.

6.3 Aktualität der Datensätze

In der zweiten Phase der Evaluation sollte nicht nur die Zuverlässigkeit des EULER-System bezüglich der Lokalisierung mathematisch relevanter Webressourcen bewertet werden, sondern sie diene hauptsächlich der Überprüfung der Aktualität des Systems. Dementsprechend waren hier die Suchfragen von Bedeutung, die auf aktuelle Sachverhalte Bezug nehmen. Die Grundlage für diese Untersuchung bildeten die mit EULER erzielten Resultate, und hier insbesondere der Vergleich der durch EULER (speziell über NetLab) lokalisierten Webressourcen mit einer direkten Abfrage des Internets mittels Metacrawler.

Eine Einordnung der ermittelten Treffer nach Relevanz war hier leider nur sehr bedingt oder gar nicht möglich. Ein solches Bestreben hätte hier zur Folge gehabt, dass jeder einzelne Link (und zuweilen wurden in einer Recherche über hundert Treffer generiert) hätte verfolgt werden müssen. Dann erst wäre die Beurteilung der Dokumente, die wiederum Schwierigkeiten mit sich brächte, möglich gewesen. Da es sich hier zumeist um sehr spezielle Fachfragen handelte, wäre eine Einstufung der aufgefundenen Dokumente nur durch einen versierten Mathematiker vorzunehmen. Zudem stellt sich - in Anbetracht der Eigenarten des Mediums Internet - die Frage, ob ein Dokument auch dann noch als Treffer zu werten ist, wenn von einer lokalisierten URL erst ein weiterer Link auf das tatsächliche Dokument verweist. Wann und wo müsste schließlich die Grenze gezogen werden?

Aufgrund dieser Umstände konnten Überprüfungen der geschilderten Art lediglich stichprobenartig durchgeführt werden. Dabei wurde die Regel befolgt, dass ein Dokument dann nicht mehr als Treffer zu werten ist, wenn von einer ermittelten URL mehr als zwei Links hätte nachgegangen werden müssen, um auf das inhaltlich relevante Dokument zu stoßen.

Lediglich in zwei Fällen (von 15 Fragen der 3. Fragekategorie: Nowadays problems in mathematics) war eine Recherche über EULER quantitativ erfolgreicher, als die Suche mittels Metacrawler. Bei fünf Anfragen konnte EULER keinen der durch Metacrawler georteten Dokumente aufspüren.¹ Jedoch ist die Aussagekraft dieser Ergebnisse sehr begrenzt, da nur die reine Menge der Trefferzahlen den Maßstab bilden konnte.

Es wurde aber dennoch (anhand der näher betrachteten Stichproben) ersichtlich, dass insbesondere die seit kurzer Zeit bestehenden Internetseiten von EULER mit abnehmender Zuverlässigkeit lokalisiert wurden. Mit der Zunahme von aktuellen Themen in die Recherche nimmt die Authentizität des über EULER gewonnene Abbildes des Internets ab. Unterstützt wird die These zusätzlich dadurch, dass verwaiste links erst nach längerer Zeit aus dem Datenbestand entfernt werden. Hierfür wurden die verwaisten URL`s, die zu Beginn der Evaluation aufgespürt werden konnten dokumentiert und über den verbleibenden Zeitraum der Untersuchung in regelmäßigen Abständen beobachtet. Es stellte sich heraus, dass „tote“ URL`s auch nach mehreren Wochen noch im System hinterlegt waren. Die mangelnde Aktualität der Daten EULER`s stellt, in Hinblick auf die Webressourcen, ein Manko dar.¹

¹ Die Eingabe der Fragen der siebten Kategorie (Information on conferences) führte zu ähnlichen Ergebnissen.

Die von EULER gebotene Ergebnisanzeige wird, insgesamt zuverlässig, dublettenbereinigt dargestellt. Bei der Anzeige von Webressourcen kommt dennoch der Umstand zum Tragen, dass mehrere links zu der selben Quelle führen. So ist es häufig möglich, dass eine URL zu einer bestimmten Homepage führt, und weitere zu Seiten der gleichen Organisation, die auch durch Links von der vorhergegangenen Homepage erreichbar sind, nochmals angezeigt werden.

Unter den von EULER angebotenen Möglichkeiten zur Spezifizierung einer Suchfrage findet sich unter der Kategorie „Resource Types“ das Hilfsmittel „Homepage“ zur direkten Lokalisierung entsprechender Seiten. Auch diese Möglichkeit erscheint durchaus nützlich, jedoch sind in erster Linie die Datensätze von NetLab geeignet, Homepages aufzuspüren. Nur werden von NetLab keinerlei entsprechende Metadaten angeboten, so dass eine Verwendung der Spezifizierung „Homepage“ nahezu immer zu einer fruchtlosen Recherche führen muss. Somit verliert diese Möglichkeit zur Einschränkung komplett an Bedeutung, da es nahezu keine Anwendungsmöglichkeiten unter diesen Voraussetzungen gibt.

6.4 Immanente Evaluation

Da nur 42% der Datensätze EULER's mittels der MSC klassifiziert sind, wurde bei der Durchführung der immanenten Untersuchung erwartet, dass eine Recherche mit den entsprechenden Verbalisierungen der Notationen zu höheren Trefferzahlen führen würde.

Während des gesamten Zeitraumes der Dateneingaben zur immanenten Untersuchung waren die Datenbanken von NetLab sowie MathNet Preprints nicht verfügbar. Eine Einbeziehung auch dieser Datenbanken hätte zur Folge gehabt (da die Datensätzen NetLab's insgesamt nicht klassifiziert sind), dass eine Veränderung der Ergebnisse zu Gunsten der Suche mittels Verbalisierung eintreten müsste.

Wie aus der entsprechenden Tabelle in Kapitel 7 ersichtlich ist, war die verbale Suche lediglich bei einem von zwanzig Fällen erfolgreicher. Bei allen anderen Eingaben führte die Recherche anhand der MSC-Notation zu höheren Trefferzahlen. Dies allein ist bereits irritierend; unter der Berücksichtigung, dass die MSC-Eingaben in einigen Fällen zu weitaus mehr Treffern führte als die verbale Eingabe, ist dieses Retrievalverhalten aber kaum erklärbar. Beispielsweise ergab die verbale Suche nach „Semimetric spaces“¹, die Recherche mit der entsprechenden MSC-Notation jedoch 17 Treffer. Eine verbale

¹ Detaillierte Informationen sind aus den nachfolgenden Tabellen, hier speziell unten den Punkten 3 (Nowadays problems in mathematics) und 7 (Information on conferences), ersichtlich.

Eingabe von „Polyadic algebras“ erbrachte 8, die MSC aber 128 Treffer usw.. Abweichungen in diesem Verhältnis können auch kaum daraus resultieren, dass bspw. NetLab zum Zeitpunkt der Eingaben über EULER nicht verfügbar war. Die separate, vergleichende Metacrawler-Recherche erbrachte zumeist höhere Trefferzahlen, als die EULER-Suche. Setzt man voraus, dass bei einer Verfügbarkeit NetLab`s die mit Metacrawler ermittelten Dokumente (und diese Annahme ist eher euphemistisch) ebenfalls lokalisiert würden, so liegt selbst die Summe der EULER- und Metacrawler-Treffer zusammengekommen häufig noch unter den Zahlen der MSC-Recherche.

Die Ergebnisse wären dadurch erklärbar, wenn die ermittelten Notationen nicht allein der vorausgegangenen Verbalisierung synonym entsprächen, sondern darüber hinaus weitere Bedeutungen aufwiesen. Eine Überprüfung der Notationen ergab jedoch eine sehr enge Übereinstimmung mit den Verbalisierungen. Auch ist diese Möglichkeit dadurch eingegrenzt, dass die verbale Suche sowohl mit dem Suchfeld „all fields“ als auch mit „keywords“ durchgeführt wurde. Beide Felder erzielen eine hohe Einbeziehung der Datenbestände, insbesondere die Verwendung von „all field“ impliziert sogar die Ermittlung von Dokumenten, deren Relevanz durchaus marginal sein kann, da die Worte bzw. Phrasen in jedem¹ der entsprechenden Dublin Core Kategorien enthalten sein können. Somit muss ein so ermitteltes Dokument nicht zwangsläufig der gewünschten Thematik vollständig entsprechen. Daher ist hier eindeutig festzuhalten, dass eine Recherche mittels MSC-Notation einer verbalen Suche vorzuziehen ist. Dies sogar in Anbetracht der Tatsache, dass nicht einmal die Hälfte der Datensätze des EULER-Systems eine MSC-Notation vorweisen. Zudem impliziert eine klassifikatorische Recherche eine hohe Präzision der Ergebnisse, setzt aber voraus, dass die jeweiligen Notationen dem Anwender auch bekannt sind. Ist dieses nicht der Fall, bietet EULER jedoch die benötigten Informationen: Sobald das Suchfeld MSC verwendet wird, öffnet sich ein Fenster, in dem das Klassifikationssystem aufgeblättert wird. Hier ist es zudem möglich, verbale Begriffe einzugeben, woraufhin die entsprechenden Notationen angezeigt werden.

¹ Diese Angabe ist insofern nicht ganz korrekt, da sich herausgestellt hat, dass das Suchfeld „all fields“ tatsächlich nicht alle Felder umfasst und somit irreführend ist. Es besteht also die Möglichkeit, dass trotz der Verwendung des All fields-Modus bestimmte Kategorien so dennoch nicht abfragbar sind. Dies zeigte sich auch durch die kuriosen Ergebnisse, die eine verbale Suche nach „Polyadic algebras“ und „Absolute planes“ erbrachte: Die Verwendung des Suchfeldes „Keywords“ produzierte höhere Trefferzahlen als die Suche mit „All fields“. Dies ist ein Umstand, der so nicht Möglich sein darf.

7. Ergebnistabellen

7.1 Tabellen zur Bewertung der Retrievalfunktionalität und Aktualität des Systems

Die nachfolgenden Tabellen zeigen die detaillierten Ergebnisse der Untersuchung.

In der ersten Spalte sind die angesprochenen Datenbanken aufgeführt. Unter „Query formulation(s)“ findet sich die jeweilige Eingabe der Suchfrage mit den jeweils verwendeten Hilfsmitteln zur Spezifizierung der Recherchen. Die dritte Spalte (Hits) zeigt zunächst oben die über EULER ermittelten Gesamttrefferzahlen. Darunter sind die Trefferzahlen gesplittet nach ihrem Ursprung, also den Quelldatenbanken, aufgeführt. In der Spalte „Source Database“ werden die Ergebnisse der direkten Abfragen in den Quelldatenbanken dargestellt. Schließlich werden unter „Relevant hits“ die Treffer separiert, die als relevant eingestuft wurden. Als „relevant“ werden hier die Gesamttreffer abzüglich der eindeutig nicht-relevanten eingeordnet. War eine solche Einstufung nicht möglich, so wird dies durch einen Strich anstelle einer Ziffer kenntlich gemacht (eine Null bedeutet jedoch, dass keiner der Treffer für „relevant“ befunden wurde). Bezugspunkte können dann lediglich die Gesamttrefferzahlen sein, die jedoch eine geringere Aussagekraft besitzen. Ein „n.a.“ steht für „not available“, d.h. die so aufgeführten Datenbanken waren zum Zeitpunkt der Eingaben über EULER nicht verfügbar. Daher konnte in diesen Fällen keine Beurteilung vorgenommen werden.

Steht anstelle einer Suchfrage mit dem verwendeten Suchwerkzeug „several queries“ heißt dies, dass die Anfragen zu keinen Treffer führten, ergänzend wurden daher weitere Anfrageformen gewählt, die aber im Einzelnen nicht aufgeführt werden, sofern letztendlich keine Treffer zustande kamen.

Unter den verwendeten Hilfsmitteln zur Recherche findet sich zuweilen bei den Abfragen von UNIFI bzw. Strasbourg „sogetto“ oder „mot clè“. Beide Terme sind mit „Keyword“ gleichzusetzen.

Das Zeichen „&“ steht für die Verwendung des Booleschen Operators „und“, „|“ für „oder“, das Zeichen „*“ verweist auf eine trunkierte Eingabe.

Aus bereits erwähnten Gründen bilden homogene Eingaben in den verschiedenen Datenbanken eine optimale Basis im Sinne der Untersuchungen. Unterschiedliche Formulierungen der Anfragen wurden nur dann verwendet, wenn keine andere Möglichkeit bestand bzw. eine gleichförmige Eingabe zu keinen Treffern führte.

1. Life and works of mathematicians from earlier periods				
Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
EULER	1. Evariste Galois any=biography bibliography&ti=galois	8		2
	MSC=01 a 70 and ti=galois	8		7
Metacrawler	"evariste galois"	0	42	
Zentralblatt	MSC=01 a 70&ti=galois	3	10	7
UNIFI	sogetto=galois	n.a.	9	0
	ti=galois	n.a.	26	1
Orsay	ti=galois	n.a.	40	2
Strasbourg	ti=galois	n.a.	25	0
EULER	1. Emmy Noether ti=emmy noether	37		10
	MSC=01 a 70&ti=noether	5		2
Metacrawler	"emmy noether"	6	87	-
Zentralblatt	MSC=01 a 70&ti=noether	8	18	6
UNIFI	sogetto=noether	1	0	0
	ti=noether	1	2	1
Orsay	ti=noether	12	7	4
Strasbourg	ti=noether	12	2	0
	motclé=noether	12	5	3
EULER	1. Carl Friedrich Gauss key=biography key=bibliography&ti=gauss	0		0
without MathDoc, Orsay/Strasbourg				
without MathDoc, UNIFI,Orsay/Strasbourg	MSC=01 a 70&ti=gauss	36		15
without MathDoc, Orsay/Strasb.	deskription=life work person&ti=gauss	18		4
Metacrawler	"carl friedrich gauss"	0	44	-

Zentralblatt	MSC=01 a 70 and ti=gauss	24	69	50
UNIFI	sogetto=gauss	-	3	0
	ti=gauss	-	6	1
Orsay	ti=gauss	-	35	5
Strasbourg	ti=gauss	-	16	3
1. Gösta Mittag-Leffler				
EULER	MSC=01 a 70&ti=mittag-leffler	3		3
without Orsay/Strasbourg, MathDoc	ti=mittag-leffler	47		6
Metacrawler	"mittag-leffler"	8	47	-
Zentralblatt	MSC=01 a 70&ti=mittag-leffler	16	6	5
UNIFI	ti=mittag-leffler	1	1	0
Orsay	ti=mittag-leffler	7	5	0
Strasbourg	ti motclé=mittag-leffler	7	0	0
1. Leopold Euler				
EULER	any=life and work&any=euler	13		10
without Orsay/Strasbourg, MathDoc	any=biography&any=bibliography&ti=euler	4		1
without Orsay/Strasbourg, MathDoc, UNIFI	MSC=01 a 70&ti=euler	14		10
Metacrawler	"leopold euler"	0	10	-
Zentralblatt	MSC=01 a 70&ti=euler	9	41	26
UNIFI	sogetto=euler	-	1	0
Orsay	ti=euler	-	40	7
Strasbourg	ti=euler	-	4	1

2. Materials on the work of contemporary mathematicians

Database	Query formulation(s)	Hits	Source Databases hits	"Relevant hits"
2. Gerd Faltings				
EULER	key ti=faltings	26		0
Metacrawler	"gerd faltings"	4	35	-
MathGuide	text ti key=faltings	0	0	0
Zentralblatt	ti=faltings	17	33	2
SUB	ti=faltings	0	10	0
UNIFI	ti key=faltings	0	0	0
Orsay	ti=faltings	0	0	0
Strasbourg	ti motclé=faltings	0	0	0
2. Friedrich Hirzebruch				
EULER	any=friedrich hirzebruch¬ Author=hirzebruch	22		11
Metacrawler	"friedrich hirzebruch"	18	30	-
MathGuide	text ti key=hirzebruch	0	0	0
Zentralblatt	bi=work&ti=hirzebruch	4	6	1
SUB	ti=hirzebruch	0	9	0
UNIFI	ti key=hirzebruch	0	0	0
Orsay	ti=hirzebruch	0	2	0
Strasbourg	ti motcle=hirzebruch	0	0	0
2. Helmut Neunzert				
EULER	ti=neunzert	5		1
Metacrawler	"helmut neunzert"	1	25	-
MathGuide	text key ti=neunzert	0	0	0
Zentralblatt	ti=neunzert	4	4	0

SUB	ti=neunzert	0	1	0
UNIFI	ti key=neunzert	0	0	0
Orsay	ti=neunzert	0	0	0
Strasbourg	ti motclé=neunzert	0	0	0
2. Andrew Wiles				
EULER	MSC=01 a*&any=wiles ti=wiles&key=wiles	7		0
		21		4
Metacrawler	"andrew wiles"	7	59	-
MathGuide	text ti key=wiles	0	0	0
Zentralblatt	ti=wiles	9	29	5
SUB	ti=wiles	0	10	1
UNIFI	ti key=wiles	0	0	0
Orsay	ti=wiles	1	0	0
Strasbourg	ti motclé=wiles	1	0	0

3. Nowadays problems in mathematics				
Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
3.1 Basic properties of double pullback graph transitions				
EULER	any=double pullback any=graph transition	2		1
Metacrawler	"basic properties"&"double pullback"&graph transition"	0	4	4
Zentralblatt	bi=double pullback&bi=graph transition	0	0	0
MathN	f=double pullback&f=graph transition	0	0	0
3.2 Equivalence of Sobolev spaces				
EULER	any=sobolev spaces&any=equivalence any=sobolev space*&any=equivalence any=sobolev space&any=equivalence	41 28 9		
Metacrawler	"sobolev spaces"&equivalence	30	20	-
Zentralblatt	bi=sobolev space*&bi=equivalence	8	94	-
MathN	f=sobolev&f=spaces&f=equivalence	0	1	1
3.3 Asymptotic freeness almost everywhere for random matrices				
EULER	any=asymptotic freeness&any=random matrices	4		2
Metacrawler	"asymptotic freeness"&"random matrices"	0	4	4
Zentralblatt	bi=asymptotic freeness&bi=random matrices	1	3	3
MathN	f=asymptotic&f=freeness&f=random&f=matrices	0	1	1
3.4 Laser cooling and stochastics				
EULER	any=laser cooling&any=stochastic*	4		2
Metacrawler	"laser cooling"&stochastic*	1	10	-
Zentralblatt	bi=laser cooling&bi=stochastic*	0	1	1
MathN	f=laser&f=cooling&f=stochastic*	0	3	3
3.5 Existence and uniqueness of weak solutions of an initial boundary value problem arising in laser dynamics				
EULER	any=weak solution*&any=initial bound*&any=Value problem&any=laser dynamic*	16		-

Metacrawler	"weak solutions"&"initial boundary value problem" &"laser Dynamics"	7	11	-
Zentralblatt	bi=weak solution&bi=initial bound*&bi=value problem& bi=laser dynamic*	0	0	0
MathN	f=uniquess&solution&boundary&laser	2	2	1
3.6 On the rareness of disturbed martingales in the class of all uniformly l 1-bounded stochastic processes				
EULER	any=rareness of disturbed martingales&any= l 1-bounded stochastic process*	3		2
Metacrawler	"rareness of disturbed martingales"& "L 1-bounded stochastic processes"	0	3	3
Zentralblatt	bi=any=rareness of disturbed martingales&any= l 1-bounded stochastic process*	0	0	0
MathNet	f=rareness&disturbed&martingales&l 1	1	0	0
3.7 Controlled algebraic k-theory				
EULER	any=controlled&any=algebraic K-theory	21		-
Metacrawler	"controlled algebraic k-theory"	12	13	13
Zentralblatt	bi=controlled&bi=algebraic k-theory	4	32	-
MathN	f=controlled&f=algebraic&f=k-theory	0	5	-
3.8 Two-primary algebraic K-theory of pointed spaces				
EULER	any=two-primary&any=algebraic k-theory &any=pointed spaces	6		4
Metacrawler	two-primary algebraic k-theory&"pointed spaces"	4	10	-
Zentralblatt	several queries	0	0	0
MathN	several queries	0	0	0
3.9 Approximation by algebraic polynomials on rectangles				
EULER	any=approximation&any=algebraic polynomial* &any=rectangle*	7		4
Metacrawler	approximation&"algebraic polynomialal*"&rectangles	4	10	-
Zentralblatt	bi=approximation&bi=algebraic polynomial*& bi=rectangle	1	3	-
MathN	several queries	0	0	-

3.10 Schottky uniformizations of bordered compact Riemann surfaces				
EULER	any=schottky uniformization*&any=riemann surface*	14	7	
Metacrawler	"schottky uniformization"&"riemann surface**"	0	10	-
Zentralblatt	bi=schottky uniformization&bi=riemann surface	5	12	-
MathN	several queries	0	0	0
3.11 Anisotropic finite elements				
EULER	any=anisotropic finite element*	17		-
Metacrawler	"anisotropic finite element"	5	20	-
Zentralblatt	bi=anisotropic finite element*	2	8	-
MathN	f=anisotropic finite element*	4	18	-
3.12 The arithmetic and geometry of algebraic cycles				
EULER	any=arithmetic&any=geometry&any=algebraic cycle*	25	-	
Metacrawler	Arimethic&"algebraic cycles"	20	10	-
Zentralblatt	bi=arithmetic&bi=geometry&bi=algebraic cycle*	3	155	-
MathN	several queries	0	0	0
3.13 Wavelets in the geosciences				
EULER	any=wavelets&any=geoscience*	21	-	
Metacrawler	all=wavelets geosciences	19	32	-
Zentralblatt	bi=wavelets&bi=geoscience*	0	0	0
MathN	several queries	0	0	0
3.14 Calabi Yau manifolds, mirror symmetry				
EULER	any=calabi yau&any=manifold*&any=mirror symmetry	47		
Metacrawler	"calabi yau manifolds"&"mirror symmetry"	10	21	-
Zentralblatt	bi=calabi yau&bi=manifold*&bi=mirror symmetry	36	109	-
MathN	several queries	0	0	-
3.15 Knizhnik Zamolodchikov equation				
EULER	any=knizhnik zamolodchikov equation	41	-	
Metacrawler	"knizhnik zamolodchikov equation"	8	20	-

Zentralblatt	bi=knizhnik zamolodchikov equation	26	81	-
MathN	f=knizhnik&f=zamolodchikov&f=equation	0	4	-

4. Popular mathematical subjects				
Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
	4. Fractals			
EULER	any=fractals&date=>1996	42	-	
Metacrawler	Fractals	4	71	-
MathGuide	text ti key=fractals	0	18	-
Zentralblatt	ti=fractals	28	458	-
SUB	rswk=fractals	4	89	-
UNIFI	key=fractals&date>1996	0	5	-
Orsay	ti=fractals	2	28	-
Strasbourg	ti=fractals	2	13	-
CWI	ti=fractals	5	21	-
MathN	ti=fractals	0	4	-
	4. Fermat (last) theorem			
EULER	any=fermat theorem	9	-	
Metacrawler	fermat theorem "fermat last theorem"	6	30	-
MathGuide	text=fermat&theorem	0	2	2
Zentralblatt	bi=fermat theorem	2	31	-
SUB	RSWK=fermat theorem fermat last theorem	0	15	-
UNIFI	key=fermat theorem	0	1	1
Orsay	ti=fermat& theorem	0	7	-
Strasbourg	motcle=fermat	0	4	0
CWI	ti=fermat theorem	1	21	-
MathN	f=fermat theorem	0	0	0
	4. Penrose tilings			

EULER	key ti=penrose tiling*	22		
Metacrawler	penrose tiling*	0	31	-
MathGuide	text=penrose&text=tilings ti=penrose&tilings key=penrose&tilings	0	0	0
Zentralblatt	ti=penrose tiling*	19	16	-
SUB	RSWK=penrose ti=penrose tiling*	0	14 0	-
UNIFI	key=penrose tiling*	0	0	-
Orsay	ti=penrose&tiling*	1	1	1
Strasbourg	ti motclé=penrose	1	1	1
CWI	ti=penrose tiling*	2	2	2
MathN	f=penrose tiling**	0	0	0
4. Four color problem				
EULER	ti=four color problem ti=four colour problem	19	-	
Metacrawler	four color problem "four colour problem"	0	5	-
MathGuide	text=four&color&problem	0	0	0
Zentralblatt	ti=four color problem ti=four colour problem	5	46	-
SUB	RSWK=four color problem	1	4	-
UNIFI	key=four color problem	1	0	-
Orsay	ti=four&color&problem	7	3	-
Strasbourg	ti motclé=four color problem	7	1	-
CWI	ti=four color problem (colour)	5	3	-
MathN	f=four color problem	0	0	0
4. Riemannsche zetafunktion				
EULER	ti=riemann* zeta*	158	-	
Metacrawler	riemann* Zeta*	6	7	-
Zentralblatt	ti=riemann* Zeta*	85	579	-
SUB	rswk=riemannsche zetafunktion	4	6	-

UNIFI	key=riemann* zeta*	5	5	-
Orsay	ti=riemann*&zeta*	26	15	-
Strasbourg	ti=riemann* zeta* motcle=riemann	26	23	-
CWI	ti=riemann zeta*	30	48	-
MathN	ti=riemann*&zeta*	0	4	-

5. Teaching material to common university courses

Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
5. Statistics				
EULER without NetLab, MathN	key=course textbook&key=statistics	44		
Metacrawler	all=statistics course	n.a.	70	-
MathGuide	text=introduction course&text=statistics	0	8	8
Zentralblatt	ti=textbook course&ti=statistics	44	69	-
SUB	ti=statistics&rswk=Lehrbuch	0	9	-
UNIFI	ti=textbook course introduction&ti=statistics	0	101	-
CWI	ti=course&statistics	0	16	-
Orsay	ti=statistics&ti=course textbook	0	40	-
5. Group theory				
EULER	key=textbook course&key=group theory	10	-	
Metacrawler	"group theory"&course	0	20	0
MathGuide	text=introduction course&group&theory	0	0	0
Zentralblatt	bi=textbook&ti=group theory	10	13	-
	msc=62-01&ti=group theory	10	55	-
SUB	ti=group theory&rswk=lehrbuch	0	1	0
UNIFI	ti=textbook course introduction&key=group theory	0	11	0
Orsay	ti=textbook course&ti=group&theory	0	40	0
Strasbourg	msc=62-01	0	42	0
CWI	ti=course&group theory	0	49	0
5. General topology				
EULER without NetLab; MathN	key=course textbook&key=general topology	2	-	
Metacrawler	phrase= general topology&course	n.a.	20	-

MathGuide	text=introduction course&text=general&topology	0	0	0
Zentralblatt	ti=textbook course&ti=general topology	2	2	-
SUB	ti=general topology&rswk=lehrbuch	0	1	-
UNIFI	ti=textbook course introduction&key=general topology	0	0	-
Orsay	ti=textbook course&general&topology	0	40	-
Strasbourg	msc=52-01	0	34	-
CWI	ti=general topology&ti=course	0	18	-
5. Linear algebra				
EULER without NetLab;MathN	key=course textbook&key=linear algebra	126	-	
Metacrawler	"linear algebra"&course	n.a.	20	-
MathGuide	text=introduction course&text=linear&algebra	1	2	2
Zentralblatt	ti=textbook course&ti=linear algebra	125	51	-
SUB	ti=linear algebra&rswk=lehrbuch	0	8	-
UNIFI	ti=textbook course introduction&key=linear algebra	0	14	-
Orsay	ti=textbook course&ti=linear&algebra	0	40	-
Strasbourg	msc=15-01	0	69	-
CWI	ti=linear algebra&ti=course	0	10	-
5. Complex analysis				
EULER	key=course textbook&ti=complex analysis	14	-	
Metacrawler	complex analysis & course	0	10	-
MathGuide	text=introduction course &text=complex&analysis	0	0	0
Zentralblatt	bi=textbook course&ti=complex analysis	14	32	-
SUB	ti=complex analysis&rswk=lehrbuch	0	6	-
UNIFI	ti=textbook course introduction&key=complex analysis	0	15	-
Orsay	ti=textbook course&ti=complex&analysis	0	40	-
Strasbourg	msc=30-01	0	54	-

CWI	ti=complex analysis&course	0	40	-
-----	----------------------------	---	----	---

6. Studies related information/organisations

Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
6. Cimpa				
EULER without SUB	any=cimpa&any=summer school	8		
Metacrawler	summer school&cima	3	13	13
MathGuide	text=cimpa	0	0	0
Zentralblatt	bi=cimpa&bi=summer school	5	5	-
6. Cime				
EULER	any=cime&any=summer school	7		
Metacrawler	"summer school"&cime	6	10	-
MathGuide	text=cime	0	0	-
Zentralblatt	bi=summer school&bi=cime	1	3	-
6. Masterstudiengang Mathematik, München				
EULER	any=münchen&any=master in mathematics	1	1	
Metacrawler	München munich&"master in mathematics"	1	11	-
MathGuide	text=master&mathematics&münchen	0	0	0
Zentralblatt	several queries	0	0	0
6. EMS				
EULER	any=ems&any=summer school	18		-
	any=ems&any=summer school*	13		-
Metacrawler	"summer school"&ems	18	20	-
MathGuide	several queries	0	0	0
Zentralblatt	several queries	0	0	0
6. Postgraduate studies				
EULER	any=mathematics&any=postgraduate studies	122	-	
Metacrawler	"postgraduate studies"&mathematics	114	25	-

MathGuide	text=postgraduate&studies	7	10	7
Zentralblatt	ti=postgraduate studies	1	2	0
6. Universität Kaiserslautern				
EULER	any=universität kaiserslautern&type=homepage	3	-	-
Metacrawler	"universität kaiserslautern"&mathematics	2	20	-
MathGuide	text=universität&kaiserslautern	1	3	1

7. Information on conferences

Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
7. Mathematisches Forschungsinstitut Oberwolfach				
EULER	any=forschungsinstitut oberwolfach	1		1
o.SUB	any=forschungsinstitut&oberwolfach&type=proceedings	22		-
Metacrawler	"forschungsinstitut oberwolfach"	1	28	-
MathGuide	text=forschungsinstitut&oberwolfach	0	0	0
7. 17th International Symposium on mathematical programming				
EULER	any=international symposium&mathematical programming&atlanta	11		-
Metacrawler	"international symposium"&"mathematical programming"	11	25	-
MathGuide	text=international&programming&symposium &mathematicians	0	0	0
7. ICM, Berlin 1998				
EULER	any=congress of mathematicians&event date=1998	1	1	1
Metacrawler	"international congress of mathematician"&1998	0	3	3
MathGuide	text=international&congress&mathematicians&berlin	0	4	3
7. 1st International congress of ISAAC				
EULER	any=1st international congress&isaac	2		2
Metacrawler	all= isaac international congress	0	29	3
MathGuide	text=international&congress&isaac	0	0	0
7. AMS Scandinavia 2000				
EULER	any=ams&mathematics meeting&odense	18		18
Metacrawler	ams&"mathematics meeting"&odense	18	24	16
MathGuide	text=ams&mathematics meeting&odense	0	0	0

8. Reference information for books and journals, OPACs

Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
8. Remmert: Funktionentheorie				
EULER	au=remmert&ti=funktionentheorie	12		10
Zentralblatt	au=remmert&ti=funktionentheorie	4	9	5
SUB	au=remmert&ti=funktionentheorie	2	5	3
UNIFI	au=remmert&ti=funktionentheorie	1	2	1
Orsay	au=remmert&ti=funktionentheorie	1	1	1
Strasbourg	au=remmert&ti=funktionentheorie	1	3	2
CWI	au=remmert&ti=funktionentheorie	1	1	1
8. Demmel: Applied numerical linear algebra				
EULER	au=demmel&ti=applied numerical algebra	4		1
Zentralblatt	au=demmel&ti=applied numerical algebra	1	1	1
SUB	au=demmel&ti=applied numerical algebra	0	1	1
UNIFI	au=demmel&ti=applied numerical algebra	1	1	1
Orsay	au=demmel&ti=applied numerical algebra	2	1	1
Strasbourg	au=demmel&ti=applied numerical algebra	2	1	1
CWI	au=demmel&ti=applied numerical algebra	0	0	0
8. Riesz/Nagy: Vorlesungen über Funktionalanalysen				
EULER	au=riesz&ti=vorlesungen	5		4
Zentralblatt	au=riesz&ti=vorlesungen	5	5	4
SUB	au=riesz&ti=vorlesungen	0	0	0
UNIFI	au=riesz&ti=vorlesungen	0	1	1
Orsay	au=riesz&ti=vorlesungen	0	0	0
Strasbourg	au=riesz&ti=vorlesungen	0	0	0

CWI	au=riesz&ti=vorlesungen	0	0	0
8. Pólya/Szedő: Aufgaben und Lehrsätze aus der Analysis				
EULER	au=polya&ti=aufgaben und lehrsätze aus der analysis	11		6
Zentralblatt	au=polya&ti=aufgaben und lehrsätze aus der analysis	6	6	6
SUB	au=polya&ti=aufgaben und lehrsätze aus der analysis	1	5	5
UNIFI	au=polya&ti=aufgaben und lehrsätze aus der analysis	0	1	1
Orsay	au=polya&ti=aufgaben und lehrsätze aus der analysis	0	4	4
Strasbourg	au=polya&ti=aufgaben und lehrsätze aus der analysis	4	4	0
CWI	au=polya&ti=aufgaben und lehrsätze aus der analysis	4	4	4
8. Lugowski/Weinert: Grundzüge der Algebra				
EULER	au=lugowski&ti=grundzüge der algebra	2	2	
Zentralblatt	au=lugowski&ti=grundzüge der algebra	2	2	2
SUB	au=lugowski&ti=grundzüge der algebra	0	1	1
UNIFI	au=lugowski&ti=grundzüge der algebra	0	0	0
Orsay	au=lugowski&ti=grundzüge der algebra	0	1	1
Strasbourg	au=lugowski&ti=grundzüge der algebra	0	0	0
CWI	au=lugowski&ti=grundzüge der algebra	0	1	1
8. Studia Mathematica				
EULER	any=studia mathematica	43		-
	ti=studia mathematica	13		10
Zentralblatt	ti=studia mathematica	2	2	2
SUB	ti=studia mathematica	7	20	-
UNIFI	ti=studia mathematica	1	2	1
Orsay	ti=studia mathematica	2	0	0
Strasbourg	ti=studia mathematica	2	0	0
CWI	ti=studia mathematica	2	1	1
8. Pentagon : a mathematics magazin for student				
EULER	ISSN=00314870	1		1
	any=pentagon&magazin	2		1

Zentralblatt	bi=pentagon&magazin	0	0	0
SUB	ti=pentagon	0	29	1
UNIFI	ti=pentagon	0	2	0
Orsay	ti=pentagon	0	0	0
Strasbourg	ti=pentagon	0	0	0
CWI	ti=pentagon	0	0	0
8. Mathematica Scandinavica				
EULER	ti=mathematica scandinavica	8		7
Zentralblatt	ti=mathematica scandinavica	0	0	0
SUB	ti=mathematica scandinavica	2	2	2
UNIFI	ti=mathematica scandinavica	1	0	0
Orsay	ti=mathematica scandinavica	3	0	0
Strasbourg	ti=mathematica scandinavica	3	0	0
CWI	ti=mathematica scandinavica	1	1	1

9. Mathematical Software				
Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
9. Mathematica				
EULER	any=mathematica&type=software	3		2
without Orsay/Strasbourg, MathDoc	any=mathematica&key=software	23		-
Metacrawler	all=mathematica software	0	53	-
MathGuide	text=mathematical&software	1	4	2
Zentralblatt	bi=software&ti=mathematica	22	57	-
SUB	ti=mathematica&RSWK=software	0	3	3
UNIFI	ti=mathematica&sogetto=programma	0	12	12
Orsay	ti=mathematica	n.a.	40	-
CWI	free: mathematica&programming	0	2	2
9. SPSS				
EULER	any=spss&type=software	0		0
without Orsay/Strasbourg, MathDoc	any=spss&key=software	6		5
Metacrawler	all=spss software	0	6	at least 20
MathGuide	text=spss	0	0	0
Zentralblatt	bi=software&ti=spss	6	8	8
SUB	RSWK=spss	0	45	45
UNIFI	ti=spss	0	1	1
Orsay	ti=spss	0	1	1
Strasbourg	ti=spss	0	0	0
CWI	ti=spss	0	8	8
9. Cinderella				
EULER	any=cinderella	28		-
without Orsay/Strasbourg,				

MathDoc	ti=cinderella	2		2
Metacrawler	all=cinderella software	29	39	>=14
MathGuide	url=cinderella	0	1	1
Zentralblatt	ti=cinderella	0	3	3
SUB	RSWK=cinderella	1	4	4
UNIFI	ti=cinderella&sogetto=programma	0	0	0
Orsay	ti=cinderella	n.a.	0	0
Strasbourg	ti=cinderella	n.a.	0	0
CWI	ti=cinderella	0	0	0

10. Images etc. of mathematical objects

Database	Query formulation(s)	Hits	Source Databases	"Relevant hits"
10. Mandelbrotmenge				
EULER without Orsay/Strasbourg, MathDoc	any=image picture&ti=mandelbrot*	15		15
Metacrawler	image: mandelbrot	13	20	20
MathGuide	text=mandelbrot	0	6	6
10. Lorenz attractor				
EULER without Orsay/Strasbourg, MathDoc	any=image picture&any=lorenz attractor*	25		19
Metacrawler	image: lorenz attractors	24	1	1
MathGuide	text=image picture&lorenz attractor*	0	0	0
10. Kleinsche Flasche				
EULER without Orsay/Strasbourg, MathDoc	any=image picture&any=klein*&any=bottle	31		29
Metacrawler	image: klein bottle	27	11	11
MathGuide	text=image picture&klein*&bottle	0	0	0
10. Schneeflockenkurve				
EULER without Orsay/Strasbourg MathDoc	any=snowflake schneeflocke* &any=image	21		20
Metacrawler	image: snowflake	21	20	-

7.2 Tabelle zur der immanenten Untersuchung

Die folgende Tabelle umfaßt die detaillierten Einzelergebnisse der immanenten Untersuchung (Notationen vs. Verbalisierung). In der ersten Spalte finden sich die verbalen Suchfragen, daneben sind die über EULER ermittelten Trefferzahlen zu sehen. Die verbalen Anfragen wurden in dem Feld „all fields“, wie auch als „keyword“ eingegeben. Es folgen die entsprechenden MSC-Notationen, in zwei Fällen steht ein Pfeil (-->) für eine Verweisung von der älteren MSC-Version auf die neue von 2000. Für die Eingabe wurde stets die 2000 er-Version verwendet. Nachfolgend werden die entsprechenden Trefferzahlen der MSC-Eingabe aufgeführt.

Die letzte Spalte beinhaltet die Trefferzahlen, die über die Metacrawle-Suche (durch verbale Eingabe) ermittelt wurden. Diese sollten lediglich eine weitere Vergleichsmöglichkeit bieten, sich aber nicht unmittelbar in der Bewertung wiederfinden.

Während der Eingabe aller Anfragen waren die Datenbanken von NetLab sowie MathN Preprints über EULER nicht verfügbar. Ein Umstand, der die Aussagekraft der Zahlen verringern mag.

Verbalisations	EULER Hits		MSC-Notation(s)	EULER Hits	Metacrawler Hits
	all fields	keywords			
Polyadic algebras	6 [!]	8 [!]	03 G 15	128	9
Umbral calculus	32	24	05 A 40	33	28
Weyl sums	19	8	11 L 15	25	19
Iwasawa theory	58	32	11 R 23	109	19
Cyclotomy	26	12	11 T 22	33	24
Excellent rings	0	0	13 F 40	2	0
Fano varieties	46	16	14 J 45	100	22
Ulm sequences	1	1	20 K 12-->20K10	66	7
Hilbert algebras	20	7	46 K 15	30	14
Dual operator algebras	8	3	47 D 27-->47L45	30	12
	2	1	51 B 15	25	14
Steiner systems	59	32	51 E 10	53	26
Absolute planes	5 [!]	6 [!]	51 F 05	24	17
Kinematic spaces	15	3	51 J 15	16	12
Nearness spaces	12	4	54 E 17	26	20
Semimetric spaces	1	1	54 E 25	17	11
Moore spaces	33	9	54 E 30	33	23
Eckmann-Hilton duality	1	1	55 P 30	3	9
Homotopy functors	5	1	55 P 65	25	12
Pfaffian systems	20	8	58 A 17	84	21

8. Schlußfolgerung

Das Projekt EJECT war auf acht Monate angelegt, die direkte Evaluation des EULER-Systems beanspruchte drei Monate. Für die Evaluation wurde eine auf die spezielle Architektur EULER's angepaßte Verfahrensweise erarbeitet. Die entwickelte Methode sollte anhand der EULER-Evaluation erprobt werden und gleichzeitig zu einer Bewertung der Retrievalfunktionalität des Systems führen. Ziel des Projektes war es zudem zu prüfen, ob die angewendete Verfahrensweise zur Evaluierung „anderer“ Subject Gateways des Internets verwendet werden kann. Somit gliedern sich die gewonnenen Erkenntnisse in zwei Punkte: Zum einen die unmittelbaren Resultate der EULER-Evaluation, komplettiert durch die dabei gemachten Erfahrungen, und zum anderen, inwieweit die angewendete Methode für die Evaluierung von Subject Gateways im allgemeinen Verwendung finden kann.

8.1 Zusammenfassende Bewertung der EULER-Retrieval-evaluation

Die Evaluation zeigte, dass EULER (mit seinem damaligen Entwicklungsstand) ein durchaus hilfreiches Werkzeug zur fachlichen Informationssuche im Sinne eines „One-stop-shop`s“ ist, dass eine effiziente Recherche ermöglicht und dabei den zu Beginn geschilderten Schwierigkeiten des Retrievals im Internet entgegenwirkt. Dabei kann mit EULER auf einen enormen Datenbestand fachlich relevanter Quellen zugegriffen werden. Dennoch weist das System gewisse Schwachstellen auf: Eine separate Suche in den einzelnen Quelldatenbanken führte insgesamt zu größeren Retrievalerfolgen, auch wenn diese weitaus mühsamer ist, als die verteilte Suche über die EULER-Suchmaske. Diese Divergenz beruht, laut Angaben der EULER-Projektpartner, auf unvollständige Datenupdates sowie auf programmtechnischen Gründen. Hierbei sind insbesondere Schwierigkeiten mit dem Z 39.50-Protokoll von Bedeutung. Auch bringt die Extraktion der Datensätze Probleme mit sich, da die Mathematik-relevanten Datensätze (dies gilt insbesondere für die OPACs der beteiligten Bibliotheken) aus den Gesamtbeständen der Anbieter herauszufiltern sind.¹ Zudem ist die Aktualität der Datensätze, was in erster Linie die Webressourcen betrifft, nicht optimal. Das Ziel des One-stop-shop kann somit nicht als vollständig erreicht angesehen werden.

Das System bietet ein reichhaltiges Angebot von Möglichkeiten und Einschränkungen zur Gestaltung der Suchanfragen, deren tatsächliche Verwendbarkeit jedoch aufgrund mangelnder Metadaten (auf die diese Bezug nehmen) teilweise stark eingeschränkt ist. Ob Möglichkeiten zur Spezifizierung, bei deren Verwendung lediglich ein sehr margi-

naler Teil aller Datensätze angesprochen wird, überhaupt angeboten werden sollen, erscheint überaus fraglich.²

Weiterhin erscheint das Design des Interfaces insgesamt nicht vollständig ausgereift, zu uneindeutig und insgesamt nicht ganz frei von Mängeln.³ In diesem Kontext erweisen sich auch die angebotenen Hilfetexte, etwa zu den jeweiligen Eingrenzungs- und Präzisierungsmöglichkeiten, als unzureichend.

Der Kern der verwendeten Evaluierungs-Methode bildeten die ermittelten Retrievalfragen und deren Eingaben in EULER und den Quelldatenbanken. Schwierigkeiten zeigten sich nicht nur bei der Ermittlung dieser Fragen, die in ihrer Breite alle wichtigen Kategorien von Anfragearten umfassen sollten, sondern insbesondere auch bei der entsprechenden homogenen Eingabeform. Für die Sammlung zweckmäßiger Fragen war die Korrespondenz mit Fachwissenschaftlern unabdingbar, da ein spezielles Fachwissen vorausgesetzt werden muss, um geeignete Kategorien und Einzelfragen zu ermitteln. Weiterhin gilt es die fachspezifische Semantik zu berücksichtigen. Die Vergleichsmöglichkeit der Retrievalergebnisse vergrößert sich, wenn die Homogenität der Eingaben gewährleistet ist. Daher ist die Homogenität der Eingaben von besonderer Bedeutung.

Für eine angemessene Berücksichtigung dieser Punkte war ein größerer Zeit- und Arbeitsaufwand erforderlich, als zunächst vermutet. Zudem zeigte sich, dass die technische Abhängigkeiten eine schwer kalkulierbare Größe bilden. Hier bestand das Hauptproblem aus der unregelmäßigen Verfügbarkeit der Systeme.

8.2 Evaluierung von Subject Gateways

Folgt man den dargestellten Kriterien zur Definition von Subject Gateways, so werden viele der Kriterien durch das EULER-System erfüllt. Allerdings weicht EULER in mindestens zwei Punkten von den vorausgesetzten Merkmalen ab: Nicht alle Datensätze des Systems haben eine intellektuelle Qualitätskontrolle durchlaufen, viele der nicht-qualitätskontrollierten Dokument sind zudem nicht mit Hilfe der MSC systematisch erschlossen. Hiervon ist ein Großteil der Internetressourcen betroffen, die wiederum einen erheblichen Teil⁴ der Gesamtzahl der Datensätze ausmachen. Außerdem ist die von

¹ Nicht alle der teilnehmenden Quelldatenbanken beinhalten rein mathematische Datensätze, daher müssen diese ermittelt und von den übrigen getrennt werden, bevor sie in das EULER-System überspielt werden können.

² Für einige Tools sind nur, vom Gesamtbestand 1 % der Datensätze mit entsprechenden Metadaten versehen.

³ Hier sei zusätzlich auf die Usability-Studie zu EULER der FH-Hamburg verwiesen: <http://www.bui.fh-hamburg.de/pers/ursula.Schulz/eulerev.htm>

⁴ Alleine die über NetLab erbrachten Internetressourcen stellen ca. 11 % des Gesamtdatenbestandes dar, diese mehr als 100.000 Dokumente sind komplett nicht klassifiziert.

EULER angebotene Browsing-Funktion keine „echte“ systematische Browsing-Funktion, sondern lediglich ein alphabetischer Index. Zusätzlich gilt es zu berücksichtigen, dass EULER ein auf rein mathematische Quellen spezialisierter Dienst ist und sich somit generell von fächerübergreifenden Angeboten unterscheidet.

Daher kann das EULER-System, im Sinne von Koch's Definitionen, nicht als Subject Gateway im engeren Sinne bezeichnet werden, sondern als eine Art Metasuchmaschine, die das Ziel eines „One-stop-shop“ verfolgt. Da aber Subject Gateways insgesamt eine noch sehr junges Hilfsmittel des Information Retrieval bilden, kann davon ausgegangen werden, dass die bestehenden Definitionen nicht als endgültig zu betrachten sind. Da sich die Einordnungen noch kaum manifestiert haben, können bestehende Definitionen lediglich als Leitfaden dienen. Spätere Korrekturen sind hier nicht auszuschließen.

Die im vorliegenden Projekt für die EULER-Architektur ausgerichtete Evaluierungsmethodik ist insgesamt nicht auf alle bestehenden Subject Gateways anwendbar, sondern nur für solche speziellen Formen, die sich durch einen EULER vergleichbaren Ansatz beschreiben lassen. Hier kommt auch der Punkt zum Tragen, dass sich Subject Gateways in der Regel - nicht wie EULER - aus den (Teil-)Datenbeständen bestehender Datenbanken zusammensetzen und somit die parallele Eingabe von Retrievalfragen in das Hauptsystem und deren Ursprungsdatenbanken oft nicht möglich ist.

Das EULER-System bietet eine neue Variante zur effizienteren fachlichen Informationssuche über das Internet. Das zugrunde liegende Konzept erscheint überaus innovativ und zweckmäßig. Mit der EULER-Engine wurde ein Prototyp geschaffen, dessen Komponenten zur Nachnutzung bereitstehen. Aufgrund der Leistungsfähigkeit EULER's mit allen Einschränkungen, die durch die Evaluation offenkundig wurden (welche teilweise bereits auch zu Nachbesserungen Anlass gaben), darf – insbesondere im Kontext wissenschaftlicher Fachgebiete - davon ausgegangen werden, dass künftig weitere Systeme entstehen werden, die eine Nähe zur Architektur EULER's aufweisen. Da die Verfahrensweise der Evaluation eine solide Methode zur Bewertung der Retrievalfunktionalität bildete und weiterhin Anregungen für weitere Bezugspunkte (Überprüfung der Aktualität, Vergleich unterschiedlicher immanenter Suchstrategien etc.) liefert, hat sich die an EULER erprobte Vorgehensweise als überaus zielgerichtet erwiesen. Der Arbeits- und Zeitaufwand ist sicherlich noch durch eine intensivere Zusammenarbeit mit den Partner (in diesem Fall das EULER-Konsortium) zu reduzieren. Die Beantwortung vieler Fragen, die sich naturgemäß erst während der Arbeit ergeben, ist für den Außenstehenden weitaus aufwendiger, als für mit dem Systemen vertrauten

Personen, da insbesondere die programmtechnischen Hintergründe im Detail kaum in den Erfahrungsbereich eines Evaluators fallen und jene für diesem auch oft nicht zugänglich sind. Dies soll jedoch nicht in Frage stellen, dass eine Evaluation durch eine Institution außerhalb der eigentlichen Projektpartner sinnvoll ist, um das notwendige Mass an Neutralität sicher zu stellen.

Eine zeitsparende Variante der Methode ließe sich dadurch erreichen, dass die Retrievalfragen in den Quelldatenbanken von den entsprechenden Instituten selber eingegeben und dokumentiert würden. Jedoch besteht hier die Gefahr, dass dies zu Lasten der Homogenität der Eingaben führen würde und so die Aussagekraft der Resultate verringert.

Übertragbar auf die Evaluation anderer Subject Gateways ist sicher die Methode, einen Satz von Suchfragen zu formulieren, der in seiner Typ-Orientierung und seiner zahlenmäßigen Ausgestaltung versucht, der realen Suchsituation möglichst zu entsprechen. Die Durchführung der Suchen und der Bewertung der Ergebnisse durch eine Person hat sich dabei bewährt. Die durch klassische Retrievaltests gewonnenen Erfahrungen haben durch diese Vorgehensweise eine Erweiterung erfahren, die durch zukünftige Untersuchungen weiter ausgebaut werden muss.